DOCUMENT RESUME

ED 344 259 CS 507 785

AUTHOR	Studdert-Kennedy, Michael, Ed.
TITLE	Speech Research, Status Report July-December 1991.
INSTITUTION	Haskins Labs., New Haven, Conn.
REPORT NO	SR-107/108
PUB DATE	Jul 91
CONTRACT	N01-HD-5-2910
NOTE	272p.; For previous report, see ED 340 053.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Collected Works - General (020)
EDRS PRICE	MF01/PC11 Plus Postage.
DESCRIPTORS	*Articulation (Speech); Communication Research; Higher Education; Infants; *Language Processing; Language Research; Models; Speech Habits; Speech Pathology; Stuttering
IDENTIFIERS	Speech Research

ABSTRACT

One of a series of semi-annual reports, this publication contains 18 articles which report the status and progress of studies on the nature of speech, instruments for its investigation, and practical applications. Articles are as follows: "The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Mode" (Catherine T. Best); "Intonational Differences between the Reduplicative Babbling of French- and English-Learning Infants" (D. H. Whalen and others); "From Babbling towards the Sound Systems of English and French: A Longitudinal Two-Case Study" (Andrea G. Levitt and Jennifer G. Aydelott Utman); "Listening with Eye and Hand: Cross-Modal Contributions to Speech Perception" (Carol A. Fowler and Dawn J. Dekle); "Recovering Tube Kinematics Using Time-Varying Acoustic Information" (R. S. McGowan); "Motor Levels of Speech Timing: Evidence from Studies of Ataxia" (Fredericka Bell-Berti and Claude Chevrie-Muller); "Implications of the Concepts Underlying Task-Dynamic Modeling on Kinematic Studies of Stuttering" (Peter J. Alfonso); "The Effect of Speech Motor Training on Stutterers' Speech Physiology" (Peter J. Alfonso and others); "Vowel Duration and Closure Duration in Voiced and Unvoiced Stops: There Are No Contrast Effects Here" (Carol A. Fowler); "P-Center Judgments Are Generally Insensitive to the Instructions Given" (D. H. Whalen and others); "Perceptual Restoration of a 'Missing' Speech Sound: Auditory Induction or Illusion?" (Bruno H. Repp); "Tongue-Twister Effects in the Silent Reading of Hearing and Deaf College Students" (Vicki L. Hanson and others); "Whole Language vs. Code Emphasis: Underlying Assumptions and Their Implications for Reading Instruction" (I. Y. Liberman and A. M. Liberman); "Depth of Phonological Recoding in Short-Term Memory" (Yi Xu); "Phonological Access of the Lexicon: Evidence from Associative Priming with Pseudohomophones" (Georgije Lukatela and M. T. Turvey); "The Contribution of Morphology to Word Recognition" (Laurie B. Feldman); "Lexical Mediation between Sight and Sound in Speechreading" (Bruno H. Repp and others); and "Leap of Faith. A Review of Derek Bickerton's 'Language and Species'" (Michael Studdert-Kennedy). A list of DTIC and ERIC numbers is appended. (RS)

CS

ED 344 259

Haskins Laboratories Status Report on
Speech Research

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

CS507785

SR-107/108
JULY-DECEMBER 1991

NEW HAVEN, CONNECTICUT

BEST COPY AVAILABLE

ED 343 529

Distribution Statement

Editor
Michael Studdert-Kennedy

Production Staff
Yvonne Manning
Zefang Wang

This publication reports the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Correspondence concerning this report should be addressed to the Editor at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut
06511-6695

Phone: (203) 865-6163 *FAX:* (203) 865-8963 *Bitnet:* HASKINS@YALEHASK
Internet: HASKINS%YALEHASK@VENUS.YCC.YALE.EDU

This Report was reproduced on recycled paper

Acknowledgment

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development

Grant HD-01994
Grant HD-21888
Contract NO1-HD-5-2910

National Institute of Health

Biomedical Research Support Grant RR-05596

National Science Foundation

Grant BNS-8820099

National Institute on Deafness and Other Communication Disorders

Grant DC 00121
Grant DC 00183
Grant DC 00403
Grant DC 00016
Grant DC 00594
Grant DC 00865
Grant DC 01147
Grant DC 00044
Grant DC 00825
Grant DC 01247

Investigators

Arthur Abramson*
Peter J. Alfonso*
Thomas Baer*
Eric Bateson*
Fredericka Bell-Berti*
Catherine T. Best*
Susan Brady*
Catherine P. Browman
Claudia Carello*
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Lois G. Dreyer*
Alice Faber
Laurie B. Feldman*
Janet Fodor*
Carol A. Fowler*
Louis Goldstein*
Carol Gracco
Vincent Gracco
Berit Halvorsen†
Vicki L. Hanson*
Katherine S. Harris*
John Hogden
Leonard Katz*
Rena Arens Krakow*
Andrea G. Levitt*
Alvin M. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist*
Virginia H. Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Patrick W. Nye
Kiyoshi Oshima‡
Kenneth Pugh*
Lawrence J. Raphael*
Bruno H. Repp
Hyla Rubin*
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Douglas Whalen

Technical/Administrative Staff

Philip Chagnon
Alice Dadourian
Michael D'Angelo
Betty J. DeLise
Lisa Fresa
Vincent Gulisano
Donald Hailey
Maura Herlihy
Raymond C. Huey*
Marion MacEachron
Yvonne Manning
Joan Martinez
William P. Scully
Richard S. Sharkany
Zefang Wang
Edward R. Wiley

Students*

Melanie Campbell
Sandra Chiang
Margaret Hall Dunn
Terri Erwin
Elizabeth Goodell
Joseph Kalinowski
Laura Koenig
Betty Kollia
Simon Levy
Salvatore Miranda
Maria Mody
Weijia Ni
Mira Peter
Nian-qi Ren
Christine Romano
Joaquin Romero
Arlyne Russo
Jeffrey Shaw
Caroline Smith
Mark Tiede
Qi Wang
Yi Xu
Elizabeth Zsiga

*Part-time
†Visiting from University of Bergen, Norway
‡Visiting from University of Tokyo, Japan

Contents

The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model
Catherine T. Best 1

Intonational Differences Between the Reduplicative Babbling of French- and English-learning Infants
D. H. Whalen, Andrea G. Levitt, and Qi Wang 31

From Babbling towards the Sound Systems of English and French: A Longitudinal Two-Case Study
Andrea G. Levitt and Jennifer G. Aydelott Utman 41

Listening with Eye and Hand: Cross-modal Contributions to Speech Perception
Carol A. Fowler and Dawn J. Dekle 63

Recovering Tube Kinematics Using Time-varying Acoustic Information
R. S. McGowan 81

Motor Levels of Speech Timing: Evidence from Studies of Ataxia
Fredericka Bell-Berti and Claude Chevrie-Muller 87

Implications of the Concepts Underlying Task-dynamic Modeling on Kinematic Studies of Stuttering
Peter J. Alfonso 93

The Effect of Speech Motor Training on Stutterers' Speech Physiology
Peter J. Alfonso, Joseph S. Kalinowski, and Robin S. Story 111

Vowel Duration and Closure Duration in Voiced and Unvoiced Stops: There Are No Contrast Effects Here
Carol A. Fowler 123

P-Center Judgments are Generally Insensitive to the Instructions Given
D. H. Whalen, André M. Cooper, and Carol A. Fowler 141

Perceptual Restoration of a "Missing" Speech Sound: Auditory Induction or Illusion?
Bruno H. Repp 147

Tongue-Twister Effects in the Silent Reading of Hearing and Deaf College Students
Vicki L. Hanson, Elizabeth W. Goodell, and Charles A. Perfetti 171

Whole Language vs. Code Emphasis: Underlying Assumptions and Their Implications for Reading Instruction
I. Y. Liberman and A. M. Liberman 181

Depth of Phonological Recoding in Short-Term Memory
Yi Xu 195

Phonological Access of the Lexicon: Evidence from Associative Priming with Pseudohomophones
Georgije Lukatela and M. T. Turvey 209

The Contribution of Morphology to Word Recognition
Laurie B. Feldman 231

Lexical Mediation Between Sight and Sound in Speechreading
Bruno H. Repp, Ram Frost, and Elizabeth Zsiga 243

Leap of Faith. A review of Derek Bickerton's *Language and Species*
Michael Studdert-Kennedy 255

Appendix 263

Haskins
Laboratories
Status Report on

Speech Research

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 1-30

The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model*

Catherine T. Best†

The Emergence of Native-Language Phonological Influences in Infants

When we hear words from an unfamiliar language spoken by a native of that language, we often have difficulty perceiving the phonetic differences among contrasting consonant (or vowel) sounds that are not distinct phonemes in our own language. Of course, we experience no difficulty with phones that are very similar to our own native phonemes. Very young infants, however, discriminate not only the segmental contrasts of their native language, but many non-native contrasts as well, as yet apparently unfettered by the phonological constraints of their language environment.[1] Moreover, young children typically come to perceive and produce just those phones that the language of their community uses, with relative ease. It is apparent, then, that the phonology of the native language comes to exert substantial influence on speech perception and production during development. The nature of the experiential effect on perception of non-native segments appears to be largely an adjustment of selective attention rather than a permanent revision of the initial state of sensory-neural mechanisms, as will be discussed later in the chapter—it is neither absolute in extent nor irremediable in adulthood, and it varies in degree among specific types of non-native contrasts and among individuals (e.g., Best, McRoberts & Sithole, 1988; Flege, 1988; MacKain, Best & Strange, 1981; Tees & Werker, 1984; Werker & Tees, 1984a; Pisoni, Logan, & Lively, in press).

When and how does the language environment come to influence the perception of phones that are not contrasting phonemes in the native sound system? And how might that developmental transition provide insight into the acquisition of the native phonological system? In particular, how do young listeners come to recognize the way in which their language organizes disparate phonetic details into phonemic categories to serve distinctly linguistic functions? These are the central issues to be examined in this chapter (see also Werker, in press). The focus will be on segmental rather than suprasegmental contrasts, particularly on consonants rather than vowels, and on non-native rather than native contrasts. That is, I will examine developmental reorganization in perception of phonetic differences among consonant sounds that do not signal a linguistic

---

The preparation of the chapter, and the original research reported here, were supported by NIH grants HD01994 to Haskins Laboratories and DC00403 to the author. The ideas presented in this chapter have gained much clarity and detail as a result of discussions with numerous colleagues and students. I would like to extend particular thanks to Carol Fowler, Janet Werker, Doug Whalen, Winifred Strange, and Michael Studdert-Kennedy for stimulating discussions and constructive criticisms of the model as I have worked on it over the past few years. I also appreciate comments on earlier versions of this chapter by Alice Faber, Alvin Liberman, Michael Studdert-Kennedy, Jim Flege, Rebecca Eilers, and the book editors. I am especially indebted to my collaborators and research assistants on the projects reported here: Gerald McRoberts, Elizabeth Goodell, Nomathemba Sithole, Deborah Wilkenfeld, Jane Womer, Stephen Luke, Rosemarie LaFleur, Shama Chaiken, Laura Klatt, Laura Hampel, Laura Miller, Leslie Turner, Pia Marinangelli, Ritaelena Mangano, Ashley Prince, Alex Feliz, Merri Rosen, Amy Wolf, Pam Speigel, Maria Poveromo, Peter Kim, Glenda Insabella, Cindy Nye, Suzanne Margiano, Diane Schrage, and Jean Silver.

distinction in the native language, but that infants in the early months can discriminate. An underlying assumption of this chapter is that developmental change in perception of non-native contrasts reflects concomitant changes in the way the child perceives the sound structure of the native language, whether at the segmental or the prosodic level.

Current findings suggest that infants begin life with language-universal abilities for discriminating segmental phonetic contrasts (i.e., are not yet perceptually constrained by the phonemic contrasts of their language environment: see footnote 1), but that by the second half-year of life listening experience with the native language has begun to influence the perception of contrasts that are nondistinctive in the native phonological system (e.g., Werker, 1989; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Lalonde, 1988; Werker & Tees, 1984b; see Werker, in press). Recent findings from my own laboratory (e.g., Best et al., 1988; Fowler, Best & McRoberts, 1990) may illuminate the means by which language-particular experience exerts its effect on perception. Specifically, the influence of the native phonological system on *adult* listeners entails the perceptual assimilation of non-native phones to the native phoneme categories with which they share the greatest similarity in phonetic characteristics. However, our findings with infants suggest that the developmental change in perception of non-native contrasts during the second half of the first year does not yet involve the mature pattern of assimilation to native phonemes. Based on these findings, I propose that the native-language influence on perception of non-native phonetic contrasts begins with the older infant's emerging recognition that native speech sounds are structured as specific, recurring constellations or patterns of coordination among phonetic-articulatory gestures (e.g., the pattern of temporal coordination between the glottal adduction gesture to begin voicing and the bilabial release gesture that characterizes English /p/). As the older infant begins to recognize these gestural constellations in ambient speech he/she may detect similar patterns in some non-native phones but may be unable to do so with others. This approach to understanding the influence of the language environment on phonemic development may also help elucidate certain aspects of phonological behavior in early speech productions, as suggested by acoustic-phonetic analyses from a case study on early word imitations to be presented in the final section of the chapter. It should be noted that while this chapter focuses on consonant contrasts, language-particular gestural constellations are assumed not to be restricted to phonemic segments, but to extend also to syllables and to units of meaning in speech (e.g., morphemes).

Language Specificity in Phonology

Languages vary in their phonological systems. Of specific interest to the present discussion, they differ in their inventories of phonemic contrasts—defined as segment-sized constellations of phonetic properties that have become *linguistically* distinctive because they are used systematically to convey differences in word meanings. For example, the lexicon of many languages including Japanese and Korean lacks the liquid consonant contrast /l/-/r/ found in English. Likewise, the English vowel contrast /ɪ/-/ɛ/ (as in <bit> - <bet>) is absent from Spanish, Italian and numerous other languages. In turn, English lacks many phonemic contrasts found in other languages, such as the ejective stops found in Ethiopian Tigrinya and elsewhere, the click consonant contrasts of African Bantu languages including Zulu, and the front rounded vowels of German, French and Swedish.

Even in cases where languages "share" a phonemic contrast, the phones involved often differ between languages in their articulatory details, i.e., in the exact phonetic realization of how those phonemes are produced. To illustrate, English and French both use the phonemic contrast /b/-/p/ to mark lexical distinctions (e.g., English <bat>-<pat>; French <bain>-<pain>). Yet in the English phoneme /b/, glottal pulsing (voicing) may begin either simultaneous with the release of the bilabial closure (the phone [b] in International Phonetic Alphabet (IPA) transcription) or slightly after the release (short-lag voiceless unaspirated [p]). In the English phoneme /p/, voicing instead begins either after a longer post-release lag, which is aspirated as an aerodynamic result of the lag in glottal adduction ([p^h]), or it begins after a shorter lag (unaspirated [p]) in certain positions (e.g., following /s/ or /ʃ/). In French, however, the phoneme /b/ is realized consistently as [b] across contexts, while /p/ is phonetically realized consistently as the short-lag voiceless unaspirated [p]. Thus, French fails to define a phonemic distinction between the phones [p]-[p^h] as English does, whereas English fails to define a phonemic contrast between [b]-[p] as French does.

By extension, languages may "share" a single phoneme which nonetheless differs phonetically between the languages, as with the phoneme /r/ found in both American English (phonetically realized as the liquid approximant phone [ɹ]) and in Spanish (realized as an alveolar tap [ɾ] or trill [r]). Furthermore, within a language there are often striking dialectal differences in the phones that typically realize a given phoneme (i.e., allophones), as with /r/, which is pronounced as [ɹ] in American English but is produced as a tapped [ɾ] or trilled [r] in Scottish English. Even within a dialect, a phoneme may have allophonic variants. These can vary either systematically, dependent on their phonetic context (context-conditioned allophones: e.g., American English /k/ realized as as [kʰ] word-initially but as [k] following /s/ or /ʃ/), or independent of context (in free variation: e.g., final /d/ may be either released or unreleased in American English). Thus, languages and dialects use but a subset of the phonetic gestures of which the human vocal tract is capable, and they differ in how they relate those articulatory details to phonemic distinctions.

These types of language-particular phonological characteristics are known to influence speech production; the degree of the experience-based effect varies with development. The phonetic details of the native language (L1) phonology are strongly ingrained in the production patterns of mature speakers who "speak like natives" not only in their choice of words but also in the accent of their speech. A corollary influence of the native phonology is that adults usually maintain an L1 accent when they learn to speak a new language (L2), and typically find it quite difficult to produce L2 with fully correct phonetic details. Yet normal young children rather quickly learn to speak the language of their community with native accents, and unlike adults, they usually can also acquire additional languages prior to 5-6 years with little or no trace of accent from L1 phonology (e.g., Brière, 1966; Flege, 1987; Flege, McCutcheon, & Smith, 1987; Oyama, 1976; Tahta, Wood, & Loewenthal, 1981; see review, Flege, 1990).

Language-specific influences on adult speech perception. But what might these observations suggest regarding the listener's *perception* of native and non-native speech sounds? Given the relative ease with which children learn to speak the ambient language(s) with appropriate phonetic detail (i.e., both L1 and L2), they certainly recognize the articulatory properties of the speech around them. Furthermore, because this occurs regardless of which particular language is being learned, we can assume that at least at some time during the developmental process of language learning, the auditory system must be capable of physiological *sensory registration* of the acoustic results of the phonetic gestures employed by natural languages.[2] But something changes developmentally with respect to the *perceptual recognition* of the organization of phonetic-articulatory details within phoneme categories. Whether or not sensory registration changes as a result of auditory exposure to particular sound patterns, perceptual recognition ability certainly does change as a result of the child's listening experience with a particular language. The latter, not the former, developmental change is of concern in this chapter, which addresses the question, "What is the nature of the developmental change in speech perception?"

One possibility is that the perceptual change results from experience with producing the sounds of L1. But a potential problem with that possibility is the paradoxical implication that correct performance can precede competence. Alternatively, the effect of the native language on perception could be independent of its influence on production. For instance, while developmental changes in ease of producing non-native sounds with correct phonetic detail might result from a history of differential articulatory practice, speech perception could remain unaffected by L1 acquisition. Although existing evidence suggests there may indeed be disparities between the improvements in production and in perception of non-native contrasts by adult L2 learners (e.g., Flege, 1988; Goto, 1971; Sheldon & Strange, 1982), there have been no direct tests of the perception-production relation in L2 acquisition during the young child's "sensitive period" for learning an L2 without an L1 accent (see Flege, 1987).

Regardless of the relation between perception and production in acquisition of L2, perceptual research has shown that the phonological characteristics of the native language do indeed influence the perceptual tendencies of mature language users. Monolingual speakers of languages with differing phoneme inventories, and/or with differing phonetic realizations of a given phoneme, show language-relevant differences in perceptual sensitivities to native versus non-native contrasts. The cross-language pattern of variations in discrimination performance on the contrasts tested have been generally consistent with the phonemic inventories of the languages studied. For example, monolingual adult speakers of English and Thai show language-appropriate differences in their perceptual boundaries between voice onset time (VOT) categories along synthetic stop consonant continua (Lisker & Abramson, 1970), monolingual Korean and Japanese speakers have difficulty discriminating the English /l/-/r/ contrast (Gillette, 1980; Goto, 1971; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975), and monolingual English speakers have difficulty discriminating the Czech voiced retroflex-alveolopalatal fricative contrast /ʐa/-/ʒa/ (Trehub, 1976) as well as several Hindi and Native American contrasts that are not used in English (Tees & Werker, 1984; Werker & Tees, 1984a).

It might be that the difficulties adults have with discriminating the pair members of many non-native contrasts reflect a permanent, absolute loss of sensory-neural sensitivity to the acoustic properties of those contrasts, or to their linguistic properties (Eimas, 1978), due to lack of environmental exposure during a critical or sensitive period of development (see Aslin & Pisoni, 1980a). However, lack of a phonemic contrast (linguistic information) need not imply lack of exposure (acoustic or phonetic data). Several factors mitigate both against an absolute lack of exposure to the non-native phonetic properties and against absolute sensory-neural loss (see also footnote 2). The absence of a given contrast does not assure that the environment is devoid of the crucial acoustic and/or phonetic properties of the phonemes involved (e.g., those patterns may be present in allophonic variants: MacKain, 1982; or in nonspeech sounds: Best et al., 1988). Moreover, perceptual recognition of phonetic organization in unfamiliar speech sounds can remain open to change even in adults learning a new language or a new dialectal accent (although the extent of malleability may be more limited than in early childhood, e.g., Flege, 1990). Discrimination of some non-native contrasts may be quite good even without training (Best et al., 1988) or with only minimal training (Werker & Tees, 1984a). Even for those non-native contrasts which are initially difficult for the listener, discrimination may become better (with some apparent limits in degree and in generalization across phoneme contexts) through extensive naturalistic conversation experience with L2 (MacKain et al., 1981; Mochizuki, 1981), or through more extensive laboratory training (e.g., McClasky, Pisoni, & Carrell, 1983; Pisoni, Aslin, Perey & Hennessy, 1982; Strange & Dittmann, 1984; Pisoni et al., in press), or under listening conditions that minimize memory demands and/or phonemic-level perceptual constraints (Werker & Logan, 1985; see also Carney, Widin & Viemeister, 1977; Pisoni & Lazarus, 1974). It is relevant here to note that there are also individual differences in the degree of difficulty listeners have with a given non-native contrast (e.g., MacKain et al., 1981; Mann, 1986; Strange & Dittmann, 1984). Finally, even when listeners have difficulty discriminating a particular non-native contrast, they have been shown to discriminate the critical acoustic features when these are isolated from the speech context (e.g., Miyawaki et al., 1975; Werker & Tees, 1984a).

These observations converge to indicate that the effect of the native language on perception of non-native contrasts is neither absolute nor permanent, and hence cannot be fully accounted for by sensory-neural mechanisms. Rather than causing a sensory-neural "loss" of sensitivity to non-native distinctions, the native language most likely promotes an adjustment of attention to language-particular, linguistic characteristics of speech signals, especially when the listener is focused at the level of phonemic information (e.g., Werker & Logan, 1985; see also footnote 2). Although the change appears to be attentional and somewhat malleable rather than strictly physiological and permanent, the facts remain that adults have initial difficulty with many (although, as I will show, not all) non-native contrasts, and that there are constraints on the extent of perceptual plasticity for those contrasts (e.g., MacKain et al., 1981; Pisoni et al., in press; Strange & Dittmann, 1984; Tees & Werker, 1984). This pattern of language-particular attunement from infancy to adulthood is consistent with the observation that perceptual learning generally involves a shift of attention away from irrelevant stimulus information, as well as increases in ability to discover and recognize functionally relevant higher-order patterns of stimulus organization (Gibson, 1966; Gibson & Gibson, 1955). How do these phonological influences of the native language on speech perception arise developmentally?

Development of phonemic influences on perception. In order to understand and speak the language of their environment, infants must come to perceive the phonetic, i.e., articulatory and/or acoustic, properties that define the phonological organization of that language. Such language-particular perceptual attunement is essential in guiding the child's own productions, if s/he is eventually to reproduce the speech patterns of other members of the language community. The universal developmental sequence and the intrinsically-motivated nature of normal language acquisition suggest that infants are well-equipped to begin making this sort of perceptual adjustment. That suggestion is also borne out by two aspects of research findings on infant speech perception, which will be discussed in more detail in the next few pages. First, studies on the discrimination of both native and non-native segmental contrasts suggest that during infants' early months, perception of segmental contrasts is largely unaffected by the language of their community (although some native-language prosodic influences may appear earlier than do segmental influences: Best, 1991a; Jusczyk, in press; Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988; cf. Best, Levitt, & McRoberts, 1991). That is, young infants' perception of segmental contrasts may reflect general prelinguistic abilities that are not yet constrained by the properties of their specific language environment (see footnote 1). This does not imply that infants possess an innate ability to discriminate all phonetic contrasts from all languages, but only that perceptual successes and failures during the first few months cut across both native and non-native segmental contrasts, revealing a pattern quite different from the native language constraints seen in adults (as discussed further below). Second, infants' phonetic perception does begin to show clear influences of the native language at least in the second half-year.

What information do infants initially perceive in speech sounds, and what are they beginning to recognize about higher-order organization in speech as they approach acquisition of their native language around the end of the first year? As they become language-users, infants must move from detecting only general information in speech (e.g., simple phonetic properties) to recognizing and producing various language-particular functional elements (e.g., words, phonemes) carried in the signal. Several key questions about this developmental transition must be addressed: Is the general information in speech that is perceptually accessible to the young infant linguistic in nature or nonlinguistic? If the infant initially perceives nonlinguistically, how does the developmental shift to perception of linguistic information take place? And finally, regardless of whether the initially detected information is linguistic or nonlinguistic, how does the infant come to recognize language-particular structural organization in the speech s/he hears? The next section will summarize the primary contemporary theoretical views and empirical investigations regarding the initial state of infant speech perception and the emergence of language-particular influences.

Development of Infant Speech Perception

Theoretical perspectives. There are three general theories of speech perception, all of which carry implications about perceptual development. One view differs from the other two in its assumptions about the nature of the information that the perceiver initially apprehends in speech, i.e., it assumes a different immediate object of perception, a different sort of perceptual primitives. That approach, referred to here as the "psychoacoustic theory," posits that the immediate object of speech perception, and hence of the infant's perceptual learning, is the *proximal stimulus*, or the raw acoustic components into which the speech signal is assumed to be decomposed by the auditory periphery. That is, that view assumes the perceptual primitives for speech perception to be an array of intrinsically meaningless, simple acoustic features, such as spectral distribution patterns, bursts of band-limited aperiodic noise, temporally-defined silent gaps, etc. into which the speech signal can be analyzed (e.g., Aslin, Pisoni, & Juszcyk, 1983; Diehl & Kluender, 1989). The psychoacoustic primitives are thus analogous in nature to the simple, two-dimensional visual features, such as edges, lines, angles and spatial frequency components, that are often described for instantaneous two-dimensional retinal images of visible patterns.

By contrast, both the current "motor theory" (Liberman & Mattingly, 1985, 1989; Mattingly & Liberman, 1988) and the "ecological theory" of speech (e.g., Best, 1984; Fowler, 1986, 1989, 1991; Fowler, Best, & McRoberts, 1990; Fowler & Dekle, 1991; Fowler & Smith, 1986; Rosenblum, 1987; Studdert-Kennedy, 1985, 1986a, 1989, 1991) argue that the immediate objects of speech perception are the *distal events* that shaped the signal (see Gibson, 1966, 1979). That is, the latter two views assume that the perceptual primitives for speech perception are the articulatory gestures of the vocal tract, i.e. the formation and release of constrictions by diverse articulators at various positions along the vocal tract (see treatments of articulatory phonological theory by Browman & Goldstein, 1986, 1989, in press). Speech signals *directly* provide articulatory-gestural information because their complex, time-varying patterns are lawfully shaped, according to the principles of acoustic physics (Fant, 1960), by the physical structure of the vocal tract and its dynamic gestures (e.g., bilabial closure, velum lowering, glottal opening). Gestural information, then, is present in the complex organization of the speech signal as it changes over time, certainly to no less an extent than pure acoustic features are present in the signal. The motor theory and the ecological view both assume that it is the gestural information that is *directly* extracted from speech signals, and that this information is *not* built up from an analysis of simple acoustic features; these views thus require no intervening mental step to "translate" raw acoustic features into gestural patterns.

The point of contention among the theories with respect to the perceptual primitives for speech perception, then, is whether the information that the perceiver extracts directly from speech is comprised of pure, simple acoustic features, or rather is comprised of dynamic articulatory patterns. Whereas the psychoacoustic approach assumes that the acoustic pressure wave is decomposed into simple, meaningless features, which serve as the immediate object of perception, the motor theory and the ecological view regard the acoustic waveform as but one of the energy media (along with the dynamic optical patterns of visible articulations, and even the haptic patterns of manually felt articulations) that are shaped by and carry information about distal vocal tract gestures, which are the immediate object of perception.

For present purposes, a brief overview will be given of some of the primary differences and similarities among these three models, as background for the view taken in the remainder of this chapter. An in-depth comparative analysis of the three models is beyond the scope of this chapter. A number of existing sources provide detailed treatments of each theoretical position, and of the debates among them; readers interested in more extended discussions of the logical and empirical grounds for each viewpoint are directed to Diehl & Kluender (1989) for an examination of the psychoacoustic approach; to Liberman and Mattingly (1985; 1988) for the presentation of the motor theory; and to Fowler (1986; see also her reply to Diehl & Kluender, 1989) and Best (1984) for discussions of the ecological account, which is based on Gibson's (1979) ecological theory of perceptual systems in general.

The three views differ with respect to how the primitives of perception are assumed to be related to the linguistic entities represented in the speech signal—e.g., phonemic segments. According to the psychoacoustic perspective, the infant must ultimately learn to associate combinations of acoustic features, which are intrinsically meaningless and nonlinguistic, with the linguistic entities of words and phrases (meaningful units), as well as of the syllables and phonemes (structural units) that may be recombined to convey different meanings (see Jusczyk, 1981; 1986, in press; Pisoni et al., in press). That is, the infant must develop auditory templates or prototypes that become paired-associates of abstract linguistic entities.

The ecological and the motor-theory perspectives assume, alternatively, that the infant must discover which particular temporal-spatial constellations of articulatory gestures are employed as specific linguistic elements (words, phonemes, etc.) in their native language, e.g. the temporal relation between bilabial closure and glottal opening at the beginning of American English words like /peak/ and /pat/. Unlike the motor theory, however, and analogous to the psychoacoustic view, the ecological view assumes that the information young infants initially detect in the speech signal is nonlinguistic. According to the ecological approach, the sort of distal articulatory information detected by the prelinguistic infant is initially devoid of linguistic relevance for them. Young infants, and indeed other animals, presumably also detect analogous distal event information in other environmental sounds and sights (see Best, 1984; Fowler et al., 1990; Studdert-Kennedy, 1986a). For example, recent studies show that young infants can recognize lawful temporal macrostructural (rhythmic) as well as microstructural (object composition) commonalities between the sights and sounds of single vs. multiple marbles being turned back and forth inside plexiglass containers (Bahrick, 1987). They come to recognize the intermodal relations on the basis of physically lawful relationships between the objects/events and the corresponding sounds, and show no evidence of learning intermodal matching when sight and sound are only arbitrarily associated

(Bahrick, 1988). Also, adults' perceptions of auditory nonspeech events involving steel balls rolling down two-part runways with differing slopes, are consistently determined by the dynamic properties of the distal events in ways that cannot be reconciled by psychoacoustic principles such as auditory contrast (Fowler, 1991).

Yet with respect to speech the human child differs from other animals because at some point in development s/he begins to discover sound-meaning correspondences between higher-order patterns of articulatory gestures and specifically linguistic functional elements, such as referents for objects, events, and people, and their interactive relationships. That is, the child discovers that mature speakers organize their articulatory gestures into systematic, recurring constellations in order to convey different meanings. These gestural constellations are the physical instantiations of the multiple, nested levels of linguistic organization in speech (phonemes, morphemes, words, phrases, sentences), all of which are specific to the language environment (see also Best, 1984). Thus, the ecological view shares with the psychoacoustic view the notion that the basis for speech perception is initially nonlinguistic in nature, sharing common ground with the perception of nonspeech sounds, and that speech perception must then shift developmentally to a linguistic basis. However, the two views obviously differ both with respect to the nature of the nonlinguistic information the infant is presumed to initially derive from speech, and with respect to the means by which the infant is presumed to make the developmental shift to a linguistic basis for speech perception.

The current motor theory (Liberman & Mattingly, 1985, 1989; Mattingly & Liberman, 1988) differs from the ecological perspective by adhering to the views that even the simple articulatory gestures that infants perceive in speech *are* linguistic in nature—specifically, they are phonetic—and that they are detected, from the outset, via a biologically-specialized language module in the human brain that is independent from the general neural mechanisms involved in the perception of all other, nonlinguistic information. On that view, the detection of distal event information is a characteristic of specialized, or closed, modules (see Fodor, 1983) but not of general, or open, modules such as the general auditory system that handles perception of other types of sounds. Here, again, the motor theory differs from the basic assumptions of both the psychoacoustic model and the ecological approach. The motor theory assumes that perception of nonspeech auditory patterns presumably proceeds more or less in the manner described by the psychoacoustic view, rather than that described by the ecological view, i.e., nonspeech auditory perception is assumed to begin with detection of the proximal acoustic properties at the auditory periphery, which does not involve detection of distal event information. According to the motor theory, the task of the uniquely human phonetic module is to relate the articulatory gestures detected in speech to the more abstract phonological units, and to translate abstract gestural structures such as words into neuromotor commands for producing specific utterances.

The view of the development of speech perception and production that is taken in this chapter follows the ecological approach. It begins with the premise that the articulatory-gestural properties of ambient speech serve as the primitives for the infant's task of learning to use speech as a tool for communicative purposes within a particular language. Analogous to the way we tend to perceive the characteristics of a physical tool (e.g., an adze) in terms of possible goal-related actions with that tool (i.e., its affordances: Gibson, 1966, 1979), learning about speech must entail a link between perception and action (speech production) in the context of affordances—outcomes that the speaker-listener perceives can be accomplished by vocal communication, such as shared games or positive emotional interactions (see also Dent, 1989)—in order for the child to become a speaker of a particular language. In this view, then, the immediate object of the infant's perception of speech is the pattern of articulatory gestures that shaped the signal; these gestures are the first and foremost properties that the infant must recognize so that s/he can come to use the vocal tract as a tool for language-specific communicative purposes.

The notion that gestural information is the basis of infant speech perception is supported by evidence that 4-6 month-old infants match visual and auditory speech patterns in bimodal perception studies (e.g., Kuhl & Meltzoff, 1982, 1984; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983), and that even younger infants show perceptual compensation for coarticulatory information in speech (Fowler et al., 1990) according to the same pattern found in adults (Mann, 1980, 1986). The latter reports considered several likely psychoacoustic explanations for the perceptual shifts they observed in differing coarticulatory contexts, including auditory contrast and critical band effects, but rejected those accounts on both logical and empirical grounds, in favor of an articulatory basis for the perceptual patterns.[3] As for the bimodal speech perception findings, the psychoacoustic account that has been offered is that the match between the optical pattern and the acoustic pattern could have been learned by association. Although no published studies have tested the association-learning account in infants, two sets of findings with adults mitigate against an associationist explanation. Specifically, synchronized but incongruent audio-visual speech stimuli often yield singular phonetic percepts that do not correspond to either the audio or the video signal, and could not have been learned by direct association (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976; cf. Massaro, 1987, 1989). Even more damaging to the associationist account are the recent findings of Fowler and Dekle (1991). In that study, subjects were tested for the McGurk-MacDonald type of cross-modal phonetic percepts under incongruous auditory-haptic presentations, in which their fingers rested against the silently moving lips of an unseen face while they were played a synchronous but phonetically-incongruent auditory token. They were also tested bimodally with incongruous auditory-orthographic presentations. Haptically-felt information about lip movements during speech is the lawful outcome of those speech movements, just as the acoustic speech signal is the lawful outcome of those same articulatory gestures, whereas orthographic symbols relate to phonetic segments by convention, i.e., by arbitrary association. Although the subjects had no prior experience with haptically felt speech, and hence could not have formed previous auditory-haptic associations, they had many years of reading experience founded explicitly on repetitive arbitrary association between phonemes and specific orthographic symbols. Nonetheless, these subjects showed cross-modal phonetic percepts akin to those found in the McGurk and MacDonald studies *only* under the auditory-haptic condition; their percepts in the auditory-orthographic condition were determined directly by the auditory stimuli and were completely unaffected by the simultaneous orthographic presentations. Thus, the results support the ecological account that speech perception is based on gestural information, and clearly run counter to the associationist account.

As discussed earlier, the distal gestural patterns of utterances are organized at multiple linguistic levels. But the ecological view assumes that these linguistic levels of organization can be detected only by perceivers who have become familiar with ambient speech and have begun to discover its affordances for conveying meaning, such that they can recognize the invariant and contrastive properties of its language-particular gestural constellations (see also Dent, 1989, for an ecological account of semantic and syntactic development). Accordingly, the emerging influence of the language environment on speech perception involves a shift from the detection of nonlinguistic information about simple gestural properties of speech, to the detection of higher-order (and functionally linguistic) coordinations among articulatory gestures. By hypothesis, this begins as the child discovers, during the final quarter of the first year, that contextually-defined references to real world objects/events (meanings) repeatedly co-occur with specific patterns of intergestural constellations in spoken words and phrases. Thus, ecological theory assumes that these emergent properties of speech gestures are themselves the linguistic entities, rather than assuming the linguistic entities to be abstract, static mental representations. Language is composed of dynamic action patterns—whether spoken, manually gestured or written—whose function is to afford speakers and listeners a means by which to communicate about actual or potential activities in which they may wish to engage (e.g., to indicate rules about a game to be played, to collaborate on a joint project).

A major appeal of the ecological approach to speech development, therefore, is its assumption that perception and production share a common metric of information—the articulatory gestures of the human vocal tract. This perception-production link is crucial to the language-learning child, who must come not only to recognize the patterns of native words across acoustically-diverse productions by widely-differing speakers, but also to produce reasonable approximations of those patterns. By the ecological account, no translation is needed between perception and production because they are informationally compatible.

The psychoacoustic approach assumes informational *in*compatibility between perception and production: the auditory percepts are specified in acoustic but not motoric terms, and the production patterns are specified in motor-control but not auditory terms, so acoustic ↔ motor translation is required. Because learning a native language requires that perception of the acoustic signal and self-production of speech be, or become, informationally compatible, the child of the psychoacoustic approach would either have to learn (construct) algorithms for linking perception and production, or else the translation routines would have to be innate. In either case, cognitive operations would have to transform acoustic and motor-control parameters to some common abstract form of information, presumably linguistic in nature, i.e., phonetic categories or features. Thus, behind the superficial acoustic "cues" or templates or prototypes must lie abstract mental representations of linguistic entities. Most existing versions of the psychoacoustic model do not address the acoustic ↔ motor translation issue. But the premises of the model would seem to mandate that the translation routines be learned associatively (e.g., Diehl & Kluender, 1989) rather than determined innately, else the psychoacoustic approach would paradoxically mirror a central tenet of the original motor theory (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The implication of associative learning for auditory-motor translation routines brings us back full circle to the problem of informational incompatibility between auditory perception and motor production, however. It is by no means a trivial matter to accomplish the necessary bootstrapping from one form to the other, especially given the problem that auditory feedback from self-produced speech could not be expected to provide guidance to motor control for the very reason that it is informationally incompatible at the outset (see also Fowler & Turvey, 1978). That is, if the auditory and motor information are incompatible, it is not clear how the child would decipher which properties of the auditory signal are, or ought to be, associated with which aspects of motor production, or how s/he would be able to evaluate whether the correct associationist inference had been formed. These basic logical problems with the implications of the psychoacoustic approach are among the primary reasons I reject that approach, and will give no further consideration to it in my subsequent discussion of early developmental changes in cross-language speech perception.

The current motor theory (Liberman & Mattingly, 1985, 1989; Mattingly & Liberman, 1988), like the ecological approach, postulates articulatory gestures as the primitives of speech perception. The motor theory requires no cognitive translation from acoustic "cues" to phonetic elements, which are directly and precognitively perceived as the distal gestures of the speaker's vocal tract articulators. However, unlike the ecological approach, the motor theory proposes that a specialized phonetic module is needed, in part, to translate the abstract gestural patterns of words, phonological elements, etc., into neuromotor commands for speech production.

The ecological approach avoids the need for translation between different forms of information either in perception or in the perception-production link, and thus has the benefit of parsimony. What the infant needs at the start is simply the general property of perceptual systems as described by Gibson (1966, 1979): that they are organized for the detection of information in stimulation about the distal objects and events that shaped the energy patterns reaching the perceiver, particularly with respect to information about the actions that those objects/events may afford the perceiver. By this general definition, human infants do not differ from other species in general approach to perceiving speech or other events. But ultimately, we do differ from other species with respect to the specific affordances that speech holds for us. We alone possess the apparatus for producing the gestures of speech (e.g., Lieberman, 1975). The few avian species that can approximate some of its acoustic patterns do so by quite different vocal mechanisms. More importantly, their imitations are not used meaningfully, and they are wholistic, failing to reorder or recombine phonetic, syllabic, or lexical sub-units to create new utterances. In fact, so far as we know, no other species systematically varies the sequence of discrete elements of meaning and/or structure in order to change the intent of their communicative messages, as the phonological and syntactic organization of human languages allow us to do. Even the apes who have been taught to use "language-like" systems of manual or visual signs to communicate with humans (e.g., Gardner & Gardner, 1973; Premack, 1971) have not mastered the grammatical functions of word order or the use of the closed class, which are both crucial characteristics of syntactic systems and hence of true language (e.g., Aitchinson, 1983; Terrace, Petitto, Sanders, & Bever, 1979). Even at the level of phonetic perception, there is recent evidence that monkeys do not organize speech categories around prototypic exemplars as human adults and infants do (Kuhl, 1991). Thus, in the context of communicative interactions, human infants move beyond other species as they discover, within the context of human social and communicative interaction, the multiple, interlocking levels of linguistic organization in speech that are carried by language-particular constellations of articulatory gestures.

Having established this theoretical background, we will turn next to a brief review of empirical findings regarding the influence of the language environment on infant speech perception, to attempt to discover what information young infants perceive in speech sounds, and how this changes as they become attuned to the native language. I will focus in particular on recent work from my laboratory on perception of non-native speech contrasts, which suggests that although the language environment has begun to influence speech perception before the end of the first year, the older infant's perception is not yet fully organized according to the phonological system of the native language that guides adults' perceptions of non-native sounds. But first we must summarize earlier findings on the general characteristics of infant speech perception, and on possible developmental influences of the language environment.

Language-particular developments in infant speech perception. Numerous studies in the past two decades have examined young infants' abilities to discriminate a wide variety of phonetic contrasts during the first half-year of life (see reviews: Jusczyk, 1985; Kuhl, 1987), although almost none of these have examined infants under 2 months. The results have indicated that, in general, 2-6 month-old infants can discriminate between-category differences for most synthetic or naturally-produced phonetic contrasts on which they have been tested. The between-category pairings that these young infants have discriminated as well as those they have had difficulty with include both native and non-native contrasts (e.g., Trehub, 1973, 1976). That is, although it would be an overstatement to claim that young infants possess an innate ability to discriminate all phonetic contrasts from all languages, nonetheless their speech perception abilities are broad, are relevant to many phonetic category distinctions, and are apparently general across languages (i.e., show cross-language similarities) rather than being biased by their specific language environment as adults' are.

The few segmental contrasts that have been reported as difficult for young infants to discriminate are consistent with the notion that general rather than language-particular abilities underlie early speech perception, because they have involved both native and non-native contrasts. In several studies, English-learning infants under 6 months have failed to discriminate certain English fricative contrasts (/f-θ/: Eilers, Wilson, & Moore, 1977; /s-z/: Eilers, 1977; Eilers & Minifie, 1975), although they have discriminated other fricative contrasts (e.g., /s/-/ʃ/: Kuhl, 1980). While some published reports have found discrimination of the difficult fricative contrasts in infants 6 months and younger (/f-θ/: Kuhl, 1980; Levitt, Juszcyk, Murray, & Carden, 1988; /s-z/: Eilers et al., 1977; Eilers, Gavin, & Oller, 1982), even the latter findings have suggested that those particular contrasts may be more difficult for young infants to discriminate than are other contrasts such as the stop place distinction /ba/-/da/.

As for non-native contrasts that may be difficult for infants under 6 months to discriminate, English-learning infants have failed to discriminate some acoustic voice-onset-time (VOT) distinctions in the range of non-native (e.g., Spanish, Thai) prevoiced-voiced stop contrasts (Eimas, Siqueland, Juszcyk, & Vigorito, 1971; Eimas, 1974). This failure was of particular interest in light of two reports that such contrasts *are* discriminated by young infants from language environments that employ prevoiced-voiced stop distinctions (Guatemalan Spanish 4-6 month olds: Lasky, Syrdal-Lasky & Klein, 1975; Kikuyu 2 month olds: Streeter, 1976a). Unfortunately, ambiguities in these studies preclude a straightforward conclusion as to whether speech perception abilities in this early developmental period show general versus language-particular constraints (see also MacKain, 1982). Two important general limitations of all these studies was their failure to directly assess for age-related changes in the perception of native vs. non-native contrasts, or to compare infants from different language environments with identical stimuli and tasks. In order to conclude that a language-particular developmental change in perception has taken place, infants from a single language environment would have to show similar responses to native and non-native contrasts at a young age, then show a language-relevant difference in response to the two contrasts at some later age. Alternatively, support for a language-particular influence would be obtained if infants from each of the two language environments showed equivalent levels of discrimination at some younger age on a contrast that is present in one but not the other language, then yielded language-relevant differences at some later age. In either case, a lack of language differences across ages would suggest that language-particular learning had not yet clearly affected perception.

There are also more specific problems in interpreting the findings. In the reports on Spanish-

and Kikuyu-learning infants, the subjects failed to discriminate the acoustic VOT contrast actually present in productions by adults of their language community, and instead discriminated some other, non-native prevoiced VOT contrast. Nor has the issue been clarified by one other report that young English-learning infants may be able to discriminate some acoustic VOT distinctions in the prevoicing range (Eimas, 1975). The latter study employed VOT differences that were much larger than the intervals used to test for category boundaries in the Spanish- and Kikuyu-learning infants, or in other regions of the acoustic VOT continuum. In addition, the use of computer-synthesized acoustic VOT continua in all these studies may pose a problem with respect to the voicing categories of the adult languages. *Articulatory* voice onset time distinctions among stop categories actually result in multidimensional acoustic differences between categories, but the synthetic continua manipulated only the timing of acoustic onset of periodicity following stop release. This obviously would be problematic if some property other than acoustic voicing onset time *per se* were the actual or primary source of perceptual information for native listeners. Indeed, even native adults' perceptual boundaries with synthetic VOT stimuli fail to correspond to the voicing categories found in Spanish (Lisker & Abramson, 1970) or Kikuyu productions (Streeter, 1976b), suggesting that acoustic VOT is not the primary perceptual information that distinguishes the prevoiced-voiced categories for them (Lisker & Abramson, 1970). And the possibility that even native voiced-voiceless stop distinctions are discriminated by infants on the basis of some other acoustic property besides timing differences in acoustic VOT is suggested by the failure of English-learning infants to discriminate synthetic /du/-/tu/ stimuli which differed only in acoustic VOT, and lacked the F1 transition cutback cue that had been confounded with VOT in other synthetic speech studies (Eilers, Morse, Gavin, & Oller, 1981).

Thus, there is only sparse, equivocal evidence of any language-specific influences on perception of consonant contrasts in infants under 6 months. The weight of empirical findings favors the view that during their first half-year infants possess only general abilities to discriminate many, though not all, consonant contrasts from both native and non-native languages. This characterization does allow that some phonetic contrasts may be easier than others for young infants to discriminate; it assumes only that such variations are not yet constrained by the infant's specific language environment. However, the possibility remains open that improvement may occur even during these early months for perception of other properties available in the native language environment, i.e., for the beginnings of some language-particular learning (for example, global prosodic properties of the native language: Mehler et al., 1988).

Other evidence indicates that the native language does begin to influence perception of phonetic contrasts during the second half-year of life. The emergence of language-particular perceptual effects during this developmental period would be consistent with general observations that infants generally start to produce their first words by the end of their first year and begin to understand words even earlier, by 8-10 months, both of which imply the development of sensitivity to the sound patterns of native words. The first such studies with older infants suggested that 6-8 month olds from different language environments may differ in their discrimination of native and non-native phonetic distinctions. Those studies examined Spanish- and English-learning infants' discrimination of synthetic versions of a prevoiced-voiced stop distinction found only in Spanish and a voiced-voiceless stop distinction found only in English (Eilers, Wilson & Moore, 1979; Eilers, Gavin & Wilson, 1979), as well as their discrimination of naturally-produced distinctions between the tapped vs. trilled /r/ ([ɾ]-[r]) found only in Spanish, the fricative voicing contrast /s/-/z/ found only in English, and the Czech fricative place contrast /ʐ/-/ʒ/ found in neither Spanish nor English (Eilers et al., 1982). The results suggested that Spanish-learning infants discriminated the Spanish voicing and /r/ contrasts, while the English-learning infants showed marginal or no discrimination of these contrasts. However, both groups discriminated the English and Czech contrasts, with the Spanish-learning infants performing no worse than the English-learning infants on the former and actually performing significantly better on the latter. The authors concluded that the language environments of the two groups of infants differentially affected their discrimination of the cross-language contrasts.

These findings suggest a possible language-particular influence on phonetic perception at 6-8 months. Although some concerns about methodological and interpretive difficulties were raised (see Aslin & Pisoni, 1980b; Jusczyk, Shea & Aslin, 1984; MacKain, 1982), the authors have rebutted many of the criticisms (Eilers, Gavin & Wilson, 1980; Eilers, Oller, Bull, & Gavin, 1984). However, some ambiguities remain nonetheless. As with the studies of younger infants, age change was not assessed. Again, one report on discrimination of non-native synthetic prevoiced/voiced stop contrasts by English-learning infants between 6-12 months (Aslin, Pisoni, Hennessy & Perey, 1981) does little to help resolve the issue. Subjects were not assessed for age changes in perception, only a very small number of the subjects who began that study completed the prevoiced/voiced tests, and even their results showed wide variations in boundary positions tested across rather wide VOT intervals (e.g., VOT differences of up to 70 ms, as compared to the 20 to 30 ms VOT intervals used in other studies with Spanish infants). An alternative explanation of language group differences at a single age in the Eilers' studies might be that the Spanish-learning infants are simply better discriminators overall than the English-learning infants for some reason other than language experience itself. Indeed, the Spanish infants discriminated the English voicing contrast as well as the English-learning infants, and they discriminated the non-native Czech contrast significantly better than did the English-learning infants. The authors suggested several additional factors, both linguistic and non-linguistic, that might account for the Spanish infants' high performance on non-Spanish contrasts: Spanish may provide more or better "phonological analogies" of those non-native contrasts than English does; the bilingual Spanish-English environment to which the Spanish infants were exposed may have aided them in discriminating the English contrast; and/or the English voicing contrast may be acoustically salient even to infants who have not been exposed to it.

More recent findings from Janet Werker's lab and from my own lab are consistent with the idea that general rather than language-particular abilities underlie discrimination of many segmental contrasts at 6-8 months. Of greater interest, however, is the related developmental finding that unequivocal language-specific changes in perception of non-native contrasts certainly *have* begun to appear around 8-10 months of age and are strong by 10-12 months. Using a version of the conditioned head-turn technique, Werker and colleagues (1981, 1984, 1988) presented English-learning infants at 6-8, 8-10, and 10-12 months with an English stop place contrast (/b/-/d/) and with the following non-native contrasts: Hindi dental-retroflex stops /d̪/-/ɖ/ and breathy voiced-voiceless dental stops /d^h/-/t^h/, Thompson Salish (Native American) ejective velar-uvular stops /k'/-/q'/. Across their several studies, the authors have used both natural CV syllables and synthetic continua, as well as both cross-sectional and longitudinal developmental designs. Yet regardless of the variation in stimuli and experimental design, the results have been remarkably consistent. At 6-8 months, the infants discriminated not only the native contrast but also all non-native contrasts, while at 10-12 months they showed significant discrimination only for the native contrast; results for the 8-10 month olds showed intermediate levels of discrimination for the non-native contrasts. For comparison, several Hindi and Salish infants tested at 10-12 months showed good discrimination of their native contrasts, the same contrasts on which the oldest English-learning infants had failed. On the basis of these findings, Werker hypothesized that the reorganization in infants' perception of non-native contrasts by 10-12 months of age reflects the emergence of the native phonological system.

Werker's findings are exciting, because they suggest that language-particular perceptual reorganization corresponds to the period during which infants are beginning to comprehend words and to establish a receptive vocabulary. It also corresponds to the period during which many infants move from producing only reduplicated babbles, in which a single syllable is repeated several times, to incorporating variations in consonant- and vowel-like elements within their multi-syllabic babbles (Oller, 1980; Stark, 1980). Moreover, at the same time, infants are making the transition from Piaget's third sensorimotor substage of secondary circular reactions to the fourth substage of means-ends differentiation. This cognitive shift suggests the possibility that while younger infants at the secondary circular reaction stage may attend to and discriminate among speech sounds because of their interesting sound properties, after shifting to the mean-ends stage infants may become more interested in speech sounds as functional means that can be directed toward communicative ends. That is, concomitant with the cognitive transition, there may be a shift toward perceiving speech sounds as members of functional linguistic categories in the infant's own language community. This could be expected to have adverse effects on perception of non-native speech sounds. Recent findings from Werker's group indicate that this cognitive transition to means-ends differentiation is indeed strongly associated with the developmental decline in perception of non-native contrasts

(Lalonde & Werker, 1990). The timing of the cognitive shift for individual infants neatly predicted their loss of discrimination for the non-native contrasts.

The Werker hypothesis about phonological influences at 10-12 months is intriguing, but suggests a different developmental pattern in perception than is provided by recent accounts of phonological development based on the productions of older, language-learning children. Specifically, it implies that the infant begins constructing his/her perceptual "map" of the native phonological system with phonemic segments as the basic building blocks. Presumably, the rapid expansion of the child's receptive vocabulary during the second year would then consist of words built up from phonemic segments. Yet researchers in child phonology have instead argued that the phonological system and phonemic contrasts are differentiated out of larger linguistic units, emerging only *after* the child has acquired a sizeable vocabulary, rather than pre-existing as the building blocks for the larger units. Recent findings from young children's speech have indicated that the earliest linguistic units are morpheme, word-, or even phrase-sized (e.g., Ferguson, 1986; Macken, 1979; Macken & Ferguson, 1983; McCune & Vihman, 1987; Menn, 1971, 1978, 1986; Waterson, 1971). From these global units, first syllables, and subsequently phonemes and phonemic contrasts, are only later differentiated (Lindblom, MacNeilage & Studdert-Kennedy, 1983), both in production (Goodell & Studdert-Kennedy, 1990; Nittrouer, Studdert-Kennedy & McGowan, 1989) and in perception (Best, 1984; Studdert-Kennedy, 1981, 1986b, 1991), most likely in response to the pressure that vocabulary growth exerts on the organization of the mental lexicon (Studdert-Kennedy, 1987; 1991). If we extend this reasoning to the emergence of language-specific influences in infant speech perception, then, we would expect language-particular reorganization in infants' perception of speech to be initiated not by the recognition of phonemic contrasts, but rather by the discovery of global patterns of gestural organization in native utterances, from which phonemes may later be differentiated.

In either case, the cross-language findings with infants raise important questions. Would the developmental pattern hold for all non-native contrasts? In particular, might there be some types of non-native contrasts that would remain discriminable because of their specific similarities to, or differences from, native contrasts? These questions can actually be traced to two underlying theoretical questions: By what means does the mature listener's phonological system affect the perception of non-native sounds? And what can the answer suggest to us about the development of the phonological system—the way in which the child moves from perceiving general information to discovering language-specific organization in the speech signal? Is the difference between the young infant and the 10-12 month old best characterized as a transition from pre-linguistic to phonological, or by some other sort of perceptual reorganization? These are the questions I have addressed in my recent research on infants' and adults' perception of various non-native contrasts, which were chosen to differ in their phonetic-articulatory relationship to categories in the phonological system of the listeners' native language. This work was sparked by a consideration of how the constraints of the native phonological system might be expected to influence the perception of non-native phonetic contrasts.[4]

A Perceptual Assimilation Model for Non-native Speech Contrasts

When presented with a speech contrast that is not employed by the native language, the mature listener is confronted with discrepancies between the properties of the non-native sounds and those of native phonemes. How do listeners respond to these discrepancies? We can dismiss the possibility that adults are unable to perceive any discrepancies, generally. For example, listeners can detect discrepancies between familiar native-accented speech and that spoken with another regional accent or even with a foreign accent (Flege, 1984). These observations indicate that mature listeners hear *discrepancies* between nonnative and native phones even though they often recognize sufficient *similarities* to familiar native phonemes to comprehend native-language utterances. What is it about the discrepancies that the listener is picking up, and how do the non-native sounds relate perceptually to native phoneme properties? The nature of the discrepancies and similarities can be viewed according to the three theoretical approaches to speech perception summarized earlier. That is, the discrepancies and similarities may be perceived in terms of either articulatory properties or acoustic properties. For reasons already discussed, this chapter takes the ecological perspective that it is primarily the evidence about articulatory gestures in the speech signal that informs the perceiver. Thus, my premise is that phonologically mature listeners perceive in non-native phones information about their gestural similarities to native phonemes. A listener will fail to detect discrepancies between native and non-native phonemes if s/he perceives the phones to be very similar in their articulatory-gestural properties to a native phoneme category. In this case, the non-native phones will be assimilated to the native phoneme category that the listener perceives to be most similar. Conversely, a listener will perceive discrepancies between native and non-native phones if s/he cannot detect a correspondence between the articulatory-gestural properties of the native and non-native phones that is even moderately acceptable. In this case, no assimilation would take place. However, assimilation is not expected to be all or none—listeners should retain some degree of sensitivity to gestural variations even within native categories (see also Best, Morrongiello, & Robson, 1981; Grieser & Kuhl, 1989; Werker & Logan, 1985), contrary to some early claims for absoluteness in categorical speech perception (e.g., Liberman et al., 1967) but consistent with subsequent evidence of above-chance within-category discrimination (e.g., Carney et al., 1977; Pisoni & Lazarus, 1974). Therefore, even as a non-native phone is assimilated to the native category perceived to be most similar, the listener often recognizes discrepancies between them (i.e., recognizes that the unfamiliar phone is less than native-like).

According to the reasoning just outlined, it follows that not all non-native contrasts should be treated alike by phonologically sophisticated listeners. Only some non-native contrasts should prove difficult for mature listeners to discriminate, while others should be easy to discriminate even without prior exposure or training. The perceptual variations should be predictable from differences in the patterns of gestural similarities and discrepancies between various non-native contrasts and the properties of native phoneme distinctions. Specifically, there are four patterns by which the two members of a given non-native contrast could be perceptually assimilated to native phonemes (Best et al., 1988):

1. The members of a non-native contrast may be gesturally similar to two different native phonemes, thereby becoming assimilated to Two Categories (TC type). For example, the Hindi retroflex stop /ɖ/ is likely to assimilate to English [d] while Hindi breathy-voiced dental stop /dʰ/ may assimilate a different English phoneme category, the voiced dental fricative [ð].

2. The non-native phones may both be assimilated equally well (or poorly) to a single native category, in which case they may be equally similar/discrepant to native exemplars of that Single Category (SC type). For example, both the Thompson Salish ejective velar /k'/ and uvular /q'/ are likely to assimilate to English [kʰ], although both will be heard as strange or discrepant from the English standard.

3. Alternatively, the non-native pair may both be assimilated to a single native category, yet one may be more similar than the other to the native phoneme, i.e., the non-native phones may show differences in Category Goodness (CG type). For example, both the Zulu voiceless aspirated velar /k/ and ejective velar /k'/ are likely to assimilate to English [kʰ], but the former should be perceived as essentially identical with English standard while the latter should be heard as quite discrepant from it.

4. Finally, the non-native sounds may be too discrepant from the gestural properties of any native categories to be assimilated into any categories of the native phonology, and should therefore be perceived as nonspeech sounds, i.e., they are Non-Assimilable (NA type). For example, the suction-produced click consonants of southern Bantu languages are unlikely to assimilate well to any English phoneme categories.

Predictions of the perceptual assimilation model. This model predicts that phonologically sophisticated listeners will show near ceiling discrimination of TC contrasts, given that the phones involved should assimilate to two different and easily discriminable native phoneme categories. These listeners should also show moderate to good discrimination of CG contrasts, which assimilate to a single native category but differ in their discrepancy from the ideal native exemplar, because they can differentiate good from less-good exemplars within the native category. However, discrimination of the CG type contrasts is not expected to reach the high levels of discrimination found for TC contrasts, because even in the native language between-category distinctions are better differentiated perceptually than are within-category variants. Mature listeners are also expected to have moderate to good discrimination of NA contrasts, but for a different reason. In this case, discrimination performance will depend on how similar the two sounds are perceived to be as nonspeech sounds.

For example, the Zulu clicks cited above may be easily discriminable if they sound like a cork popping vs. fingers snapping, or else they may be only moderately discriminable if they sound like two different finger snaps. Different CG and NA contrasts may vary in discriminability due to variations in the degree of similarity, respectively, in their phonetic-articulatory properties or in their auditory properties. Finally, mature listeners are expected to show poor discrimination of SC contrasts, because the two phones assimilate to a single native phoneme category but are equally discrepant from the standard exemplar of that category.

Thus, the discrimination performance pattern for adults should be, from highest performance to lowest: TC > (NA<=>CG) > SC. This prediction assumes strong phonological influence from the native language, and is precisely the pattern of performance we have obtained with adult listeners across several experiments with non-native speech contrasts, as I will summarize. It should be noted that CG and SC contrasts fall at different ends of a single dimension, in that both involve assimilation of a non-native phone pair to a single native category. In the case of CG contrasts, discrimination is aided by the listener's recognition that one non-native phone is more discrepant from the standard native exemplar than is the other. A corollary of this principle is that various CG contrasts may differ in the degree of discrepancy between the two non-native phones, and hence may vary in degree of discriminability. If the discrepancy from the native prototype is large for one non-native item but very small for the other (a strong CG difference), discrimination will be better than if there is only a small difference in discrepancy between the two non-native items (a weak CG difference). At the extreme of this dimension of assimilability, both members of the non-native contrast are equally discrepant from native category exemplars, in which case we have a SC contrast with poor discriminability. Also note that most of the earlier-studied non-native contrasts that have proven difficult for adults and older infants to discriminate fit the definition of SC contrasts or of weak CG contrasts, which could account for the listeners' difficulties. In a few previous reports, adults have fared better with some non-native contrasts even with little or no training, as in English-speaking listeners' discrimination of the Hindi voiceless aspirated /t^h/ vs. breathy voiced /d^h/ stops (Werker & Tees, 1984a), and Kikuyu-speaking listeners' discrimination of the English voiced-voiceless stop distinction (Streeter, 1976b). The latter cases fit the definition of a TC contrast and a strong CG contrast, respectively. Finally, it should be noted that NA contrasts, like CG contrasts, theoretically may vary in degree of discriminability, which will in these cases be determined by variations in salience of the *auditory* differences between pair members (e.g., Burnham, 1986). Auditory rather than phonetic-articulatory differences should determine discrimination of NA contrasts because phonologically sophisticated listeners are expected to perceive them as nonspeech sounds; that is, as nonlinguistic mouth sounds or perhaps as sounds produced by similar nonvocal events.

The predictions for phonologically sophisticated adult listeners are clear. The expectations for young infants under 6-8 months are likewise clear, but different from those for adults. Specifically, young infants' discrimination performance should not differ in a phonologically consistent way according to the four assimilation types, but rather should be good for most native and non-native contrasts. To the extent that these young infants may show different discrimination levels for various contrasts, these variations should not follow the pattern described for adults but should instead be related to nonlinguistic differences in the complexity or salience (to the young infant) of the phonetic-articulatory distinctions involved.

On logical and/or theoretical grounds, however, there are several possible outcomes for the older 10-12 month old infants, who have shown clear evidence in previous studies of a dramatic reorganization in perception of non-native contrasts. Their pattern of performance on the four assimilation types should provide insight into the nature of that perceptual reorganization. One logical possibility is that advances in the infant's general cognitive/memory abilities may affect their responses to stimulus familiarity/novelty, perhaps leading to a simple heuristic in which sounds (including speech sounds) that occur in the ambient environment are recognized as familiar, while those that do not occur are unfamiliar and therefore pose perceptual difficulties (but see earlier discussion, and also MacKain, 1982 for problems with the underlying assumptions of this reasoning with respect to the language environment). On this account, at least as regards speech perception, older infants on the verge of language acquisition should become less sensitive to and/or interested in contrasts between phones that are both absent from their environment (i.e., both unfamiliar). By definition, this description would fit TC, NA, and SC contrasts, and at least weak CG contrasts. Older infants should therefore show poor discrimination of all non-native contrasts except perhaps the strong CG type, which contrasts a familiar native-like phone against an unfamiliar one. This set of predictions I will call the general **familiarity hypothesis**.

Alternatively, however, the perceptual shift may be specifically linguistic in nature, rather than simply being an instance of a general, language-independent change in response to unfamiliar stimuli. There are several potential patterns by which a linguistically-based reorganization might occur. If the perceptual shift by 10-12 months is a reflection that perception has shown a stage-like shift to becoming rule-governed by the phonological system of the native language, then these older infants should show the same pattern of phonologically-based discrimination performance as the adults of their language community. That is, they should show excellent discrimination of TC contrasts, good-to-moderate discrimination of CG and NA contrasts, and poor discrimination of SC contrasts only. I will refer to this view as the **strong phonological hypothesis**, to reflect its prediction of the infant's stage-like emergence into a higher, linguistic level of perceptual organization that is governed by the native phonological system. Note that this approach entails the infant's recognition of the linguistic function of phonemic contrasts and other phonological rules such as allophonic distributional constraints, which are problematic assumptions (see MacKain, 1982, and discussion in the section *Language-particular developments in infant speech perception*).

Two other possible reorganization patterns seemed more likely, each of which would indicate a different path for the infant's developing recognition of native phones and phonemic contrasts. One of these possibilities is that the older infants' perception is organized according to phonemic contrasts, but that their recognition of the patterns of coordination among phonetic details for individual native phone categories is still under-differentiated. That is, they would be expected to discriminate clear between-category contrasts, but may show greater acceptance of "deviant" tokens *within* a given phone category than adults do. That is, they may show lower within-category discrimination of the differences between "good" and "poor" exemplars, suggesting less refined mapping of the prototype space within the category (but see Grieser & Kuhl, 1989). This view could be referred to as the **phonemic contrast hypothesis**. Once again, the hypothesis entails the problematic assumption that infants recognize the linguistic function of phonemic contrasts. In this scenario, the older infants would discriminate TC contrasts, but because of their under-differentiated recognition of the coordinated phonetic details within individual native phone categories, they would have difficulty discriminating CG contrasts, which entail within-category distinctions between good and less-good exemplars of a single phoneme. SC contrasts would likewise become difficult to discriminate as variants of a single native phone, although NA contrasts would remain discriminable as non-phones (i.e., as nonspeech). Thus, this view differs from the familiarity hypothesis by predicting good discrimination of TC contrasts, and it differs from both that hypothesis and the strong phonological hypothesis by predicting poor discrimination of CG contrasts.

According to the final hypothesis, 10-12 month olds' perception may *not* be organized around pairwise phonemic contrasts as functional linguistic oppositions, but rather may focus on the recognition of the patterns of gestural coordination that identify members within a given native category—a **category recognition hypothesis**. Note that the categories the infant comes to recognize need not, in fact, be confined to phonetic *segments*, but may also include larger gestural units such as syllables, words, etc. This last hypothesis, then, may be most compatible not only with the ecological view espoused in this chapter, but also with the previously-mentioned arguments that the child's earliest linguistic units are words, morphemes, and sometimes phrases (e.g., Ferguson, 1986; Macken, 1979; Macken & Ferguson, 1983; Menn, 1971, 1978, 1986; Waterson, 1971), and that segmental phonology—phonemes whose boundaries are defined within a system of phonological contrasts—emerges only later by differentiation from these larger units (Lindblom, MacNeilage & Studdert-Kennedy, 1983) in response to the pressure that vocabulary growth exerts on the organization of the child's lexicon (Studdert-Kennedy, 1987; 1991). In other words, functional phonemic contrasts *between* phone categories are hierarchically more complex and abstract than are the coordinated gestural patterns that define category membership of a given utterance or phone. Nonetheless, the infant's recognition that two gestural coordination patterns differ from one another would be expected to lead to good discrimination of the two phones, but on the basis of recognizing either that each phone is a clear member of a different phone category (i.e., the case of native phonemic contrasts) or that one phone is a clear member of a given category while the other is not a good member of that category (i.e., the case of a non-native strong CG contrast). By this final hypothesis, the older infants should have difficulty discriminating SC contrasts, as well as TC contrasts for which the non-native phones are both unrecognizable to the infant as any native gestural coordination patterns. That is, while older infants could be expected to recognize many non-native sounds as speech sounds because they can detect in them some of the general articulatory properties found in native speech, nonetheless they should find it difficult to detect sufficient similarity between some non-native gestural coordination patterns and the patterns they have discovered in specific native phones. This should make it difficult for older infants to discriminate not only the unfamiliar gestural coordinations seen in SC contrasts, but also those TC phones whose gestural patterns deviate in many ways from even the most similar native phones. On the other hand, they should discriminate at least some CG contrasts as involving a native-like gestural constellation opposed to an unfamiliar gestural constellation, although they may not discriminate even those contrasts as well as adults do. However, they should perceive NA phones as non-speech sounds (non-phones) because in them they would fail to detect even any global gestural similarities to native phones; hence, they should discriminate NA contrasts on the basis of their nonspeech properties. Thus, the category recognition hypothesis differs from both the strong phonological hypothesis and the phonemic contrast hypothesis by predicting poor discrimination of some TC contrasts; from the phonemic contrast hypothesis also by predicting good discrimination of some CG contrasts; and from the general familiarity hypothesis by predicting good discrimination of NA contrasts and of some TC contrasts.

Empirical investigations of perceptual assimilation. In the first test of the model's predictions, we assessed English-speaking adults' and infants' discrimination of a NA non-native contrast, the Zulu apical vs. lateral click contrast /ʇa/-/ʖa/ (Best et al., 1988). This non-native pair was expected to be non-assimilable to any English phonemes because the suction-release gesture used in them is not employed in any English phones, nor is it reasonably similar to any English gestural maneuvers in the way that implosive stops are gesturally similar to plosive voiced stops, or the way that ejective stops are similar to voiceless stops, except for variations in their laryngeal maneuvers. Moreover, the asymmetrical lingual release of the lateral click is not found in any English phonemes. The apical and lateral clicks sometimes do appear in isolation (i.e., without vowels) as nonspeech "mouth sounds" in our culture, the former appearing as "tsk-tsk" sounds that indicate frustration or disapproval, the latter as a "chucking" sound used to indicate approval or to urge a horse along. These nonspeech occurrences may reinforce the American listener's tendency to perceive the Zulu clicks as nonspeech sounds.

We began by testing American adults' discrimination of the 18 minimal-pair contrasts among the 9 nonnasalized Zulu clicks (apical, lateral and palatoalveolar places of articulation crossed with prevoiced, voiced, or voiceless-aspirated manner) in click + /a/ syllables. The AXB discrimination task employed multiple natural tokens of each category (X was a physically different token from both A and B), and thus depended on some degree of perceptual constancy for successful completion. There was no training on the click contrasts, and just a few practice trials to orient subjects to the task. Even with this minimal exposure to Zulu clicks, the American adults discriminated all contrasts well above chance, as predicted, showing between 85-95% correct discrimination for all pairings except one. Moreover, the subjects' responses on a posttest questionnaire revealed that, as predicted, they had indeed perceived the clicks as nonspeech sounds made by release of tongue suction (e.g., "tongue clucking") or as other sounds resulting from abrupt pressure changes (e.g., "cork popping"), rather than perceiving them as phonemic segments. Their lowest performance was 80% correct discrimination with the apical vs. lateral voiced (short-lag VOT) pair, so this contrast was chosen for further testing with adults and infants.

To eliminate the most obvious acoustic difference between the two click categories, we equated the amplitudes of the clicks, and verified that they were still acceptable category exemplars in standard identification and discrimination tests with six Zulu listeners. A second AXB discrimination test with a new group of American adults found discrimination to be virtually as good as before the stimulus manipulations, around 78% correct. We then tested English-learning 6-8, 8-10, and 10-12 month olds (the ages examined by Werker and colleagues, 1981, 1984b, 1989), 10 at each age, for discrimination of this click contrast and of a control English contrast (/ba/-/da/), with test order counterbalanced in each age group. For comparison, we tested another group of adults using the infant procedure, and extended the test to even older infants at 12-14 months. In each test, the subject was conditioned to fixate on a colorful slide to hear repetitions of the multiple natural tokens of the habituation syllable, which terminated whenever the infant looked away from the slide. Following a significant decline in fixation times below a criterion habituation level for two consecutive trials, the audio presentations were shifted to the test stimulus for that non-native contrast. To assess for discrimination, mean looking times were computed for the two trials immediately preceding the stimulus shift (habituation level) and for the first two post-shift trials (response recovery).

The results of the infant study upheld the prediction that discrimination would remain high across all ages. The younger infants discriminated the category change and, moreover, there was no evidence of a decline in discrimination among the older infants or adults (in fact, adults showed better discrimination than the infants in this habituation task—see also Eilers, Wilson & Moore, 1979). Thus, the findings differ from those reported by Werker et al. (1981, 1984b, 1989) in that although the click contrast is a non-native distinction, older infants do *not* lose sensitivity to it. Our results are compatible with the claim that "loss" of discrimination for non-native contrasts is not absolute and across the board, but rather is due to differences in perceptual assimilation. This finding is at odds with the general familiarity hypothesis about the nature of the developmental change in perception of non-native contrasts, but is still compatible with the predictions of each of the other three hypotheses.

That first study had employed a different experimental procedure than Werker had used, however, so the possibility remained that the difference between her results and ours could be traced to methodological factors rather than to assimilation differences between the Zulu clicks and the non-native phonetic contrasts she had used. Therefore, in a second study we replicated both our findings with the Zulu click contrast and Werker's findings with her Salish ejectives, an SC contrast (Best & McRoberts, 1989). We used our visual fixation procedure with new groups of 6-8 vs. 10-12 month olds, 12 per age. Each infant was tested on both of the non-native contrasts as well as on the English control contrast. We again found that both age groups discriminated the NA Zulu contrast and the English control, whereas only the younger infants discriminated the SC Salish ejectives. Thus, the developmental difference between the two non-native contrasts could be attributed to differences in perceptual reorganization for those types of contrast, and not simply to methodological factors. This finding is again at odds with the general familiarity hypothesis, but still fails to differentiate among the three linguistic hypotheses.

The next step, then, was to compare discrimination of other non-native assimilation types. For this purpose, we examined three additional contrasts from Zulu. The TC contrast was a lateral fricative voicing distinction /ɬɛ/-/ɮɛ/, produced with the tongue in essentially the position used for English /l/ but with a greater degree of constriction along the sides of the tongue. For adult American listeners, the voiceless lateral fricative would be expected to assimilate to the English voiceless coronal fricatives /s/, /ʃ/ or /θ/, perhaps heard with a (devoiced) subsequent /l/ , due to its /l/-like positioning of the tongue tip/blade co-occurring with fricative manner and voicelessness (see Ladefoged, 1981).[5] The voiced lateral fricative would be expected to assimilate to the English voiced fricatives /z/, /ʒ/ or /ð/, and/or the approximant /l/. The CG contrast was a velar voiceless vs. ejective stop distinction /ka/-/k'a/. The Zulu /k/ is virtually identical to English /k/ (both [k^{h}]). But the ejective /k'/ involves a non-native glottal gesture (rapid upward movement of the glottis during complete glottal closure) that should lead to its assimilation as a clearly less than ideal exemplar of English /k/; thus, this stimulus pair constituted a strong CG contrast. The remaining contrast was a plosive vs. implosive bilabial stop distinction /bu/-/ɓu/, originally chosen to represent an SC contrast. However, further consideration of the articulatory properties of these phones suggested that it was actually a weak CG contrast. Zulu /b/ is essentially like English /b/, but the downward movement of adducted and vibrating vocal folds for the implosive /ɓ/ differs primarily in degree from the laryngeal movement involved in English /b/. According to the strong phonological hypothesis for mature listeners, the English-speaking adults would be expected to discriminate the TC contrast nearly perfectly, the strong CG contrast somewhat less well, and the weak CG contrast most poorly but above chance.

In the first part of the study, 25 monolingual English-speaking adults completed separate AXB category-identity discrimination tests on the three contrasts (as in Best et al., 1988), each test again composed from multiple natural tokens. As predicted, adults discriminated the TC contrast with near ceiling performance levels (~ 96% correct). They did slightly but significantly less well with the Zulu CG contrast (~ 88%), and discrimination performance was substantially and significantly lower on the SC contrast, although it was nonetheless above chance level (~ 66%). The subjects' posttest questionnaire descriptions of the Zulu sounds indicated that nearly all subjects assimilated the TC contrast to two different English phonemes or phoneme clusters, although there was variability in which native phonemes (or clusters) were named (s, sh, thl, etc. versus z, zh, zhl, or l), compatible with the gestural similarities and discrepancies from various English phonemes. Most heard the strong CG contrast as a normal /k/ versus a clearly deviant /k/ (e.g., "choked" or "coughed"). And as expected, the majority heard the weak CG items as exemplars of English /b/; only some of these subjects could identify a difference, in which they generally characterized one /b/ as "murmured" or "swallowed." Thus, consistent with the perceptual assimilation model, adults appear to assimilate non-native contrasts to the closest native phoneme categories, apparently on the basis of articulatory similarities and discrepancies. Furthermore, the discrimination performance pattern precisely mirrors the predictions of the strong phonological hypothesis, as we expected for phonologically sophisticated listeners. We recently verified near-ceiling performance with another TC contrast, the Ethiopian ejective stop distinction /p'ɛ/-/t'ɛ/, which this time was assimilated virtually unanimously to English /p/-/t/, as expected from the straightforward gestural correspondence between the non-native and native sounds (Best, 1990).

In part two of the second Zulu study, English-learning 6-8 and 10-12 month-olds, 14 per age, completed visual-fixation habituation tests for each of the three contrasts (Best et al., 1990). The predictions from all four developmental reorganization hypotheses, as summarized earlier, were that 6-8 month old infants would discriminate all three contrasts. The three linguistic hypotheses offered different predictions for the 10-12 month olds, as outlined previously. Analyses of the data for the 6-8 month olds indicated significant discrimination across all three contrasts, consistent with predictions and with earlier findings. However, discrimination of the TC contrast by itself was marginal, while discrimination was significant for each of the CG contrasts. Contrary to both the younger infants and the adults, however, the 10-12 month olds showed only marginal discrimination (p < .06) across these three Zulu contrasts, although this age had clearly discriminated the NA Zulu clicks in our two previous studies. Moreover, the contrast on which they showed the poorest performance—actually a small *decline* in fixation times at the stimulus shift—was the TC contrast that had proven easiest for the adults to discriminate. This age showed marginal discrimination of the strong CG contrast, and nonsignificant discrimination of the weak CG contrast. Although the largest recovery in mean fixation time for either age was associated with the 10-12 month-olds' response to strong CG contrast, the statistical effect was hampered by high intersubject variability. This pattern of a high mean recovery paired with high intersubject variability suggests that some older infants may have detected the category change while others utterly failed to. Planned comparisons on the 10-12 month-olds' discrimination pattern failed to support the strong phonological hypothesis (TC > strong CG > weak CG discrimination), as well as the phonemic contrast hypothesis (TC discrimination only), but did offer marginal support (p < .08) for a pattern compatible with either the category recognition hypothesis or the general familiarity hypothesis (strong CG discrimination only).

Given that the two previous studies were consistent with the three linguistic hypotheses but not with the familiarity hypothesis, then, the full set of findings is most compatible with the category recognition account of perceptual reorganization at this age. But the category recognition hypothesis states that at least some TC contrasts should be discriminable to this age group. Why did this TC contrast pose such difficulty for the older babies? Two observations suggest a possible answer, although further research is needed to confirm it. First, recall that the gestural properties of the TC phones (lateral voiced and voiceless fricatives) do not provide a close match to any singular English phonemes. These Zulu phones involve a lingual gesture similar, but not identical, to that found in our lateral approximant /l/, yet they also involve lingual constrictions generally similar, but not identical, to a variety of English fricatives (/s/, /ʃ/, /θ/ and their voiced counterparts). Thus, it would be understandable if the 10-12 month old, who is only beginning to recognize the gestural coordination patterns found in English phones, has difficulty recognizing any clear similarities between the Zulu phones and particular English categories. Indeed, even the adults were quite variable in the exact patterns by which they assimilated the phones in this TC contrast to specific English phonemes. Second, the older, and even the younger, infants' difficulty with the lateral fricatives does not appear to be a general problem either with non-native TC contrasts. We recently obtained evidence that, like adults, both 6-8 and 10-12 month olds *can* discriminate an Ethiopian ejective TC contrast /p'/-/t'/, which bears a straightforward gestural relation to a single English distinction, /p/-/t/. Moreover, both ages also discriminated the English fricative voicing distinction /s/-/z/, which is similar to the Zulu TC fricative distinction that the older infants failed to discriminate (Best, 1991b). It is interesting to note, however, that the older infants showed marginally lower discrimination than the younger infants on the English /s/-/z/ contrast, which involves only a relative shift in voicing onset during an ongoing frication. Thus, it appears likely that the reason the older infants had difficulty with the Zulu fricative TC contrast, but not with the Ethiopian ejective contrast, was because the gestural properties of the ejectives relate more straightforwardly to English stops than the Zulu fricatives relate to any clear English categories.

Thus, the pattern of findings across studies for the older infants are at odds not only with the general familiarity hypothesis but also with both the strong phonological hypothesis and the phonemic contrast hypothesis. The performance of the 10-12 month olds on NA, SC, CG and TC contrasts across studies is most compatible with the category recognition hypothesis. These results carry strong implications about the nature of developmental change in infants' perception of non-native speech sounds, and by extension offer insights about the development of the native phonological system. They suggest that by at least 10-12 months of age, infants have begun to discover the gestural coordination patterns that identify categories roughly corresponding to phones in their native language. The findings also indicate that at least some of their categories still may not be as well differentiated as adults' and may not be as strongly organized according to the pairwise linguistic *contrasts* of the native phonological system. More specifically, adults appear to assimilate non-native phonemes to categories within the native language's system of phonemic contrasts, yielding near-ceiling discrimination performance on both of the tested non-native TC assimilation contrasts. The older infants, however, and even the younger infants had difficulty discriminating the TC fricative contrast for which adults showed good discrimination but variable patterns of assimilation, while neither of the infant age groups had any difficulty discriminating another TC ejective stop contrast for which adults were in perfect agreement about assimilation. Moreover, a study recently completed in my lab (Insabella, 1990 Insabella & Best, 1990) confirms that the Zulu TC fricative voicing contrast continues to pose difficulty even at 4 years of age, by which time the strong CG contrast (Zulu /k/-/k'/) is discriminated consistently.

The pattern of findings with the infants, and the extension to 4 year olds, suggests that perception of non-native speech contrasts in relation to the system of phonemic contrasts in the native language is a relatively late achievement that probably rests on a solid foundation of knowledge about the coordinated phonetic patterns of individual native phones. The 10-12 month old infants' marginal discrimination of the strong CG contrast, compared with the greater difficulty both ages showed on the TC lateral fricative contrast, runs counter to the predictions of the phonemic contrast hypothesis. The initial stage of language-specific influence on perception of phonetic segments appears to involve the emerging recognition of the coordination of phonetic-gestural details *within* individual phone categories, rather than recognition of the phonetic distinctions that specify more abstract and linguistically functioning phonemic contrasts *between* categories. In keeping with the ecological theoretical perspective outlined earlier in the chapter, I suggest that the basis for infants' recognition of the language-specific properties of native (and non-native) phones is the detection of evidence about the constellation of coordinated articulatory gestures that are associated with specific phones in the native language.

Drawing from these findings, I suggest the following developmental progression in language-specific perceptual learning about speech:

1) Young infants initially perceive simple nonlinguistic (articulatory and/or acoustic) distinctions in speech sound contrasts, and this ability is not yet influenced by their language environment.

2) By at least 10-12 months, infants have begun to discover certain gestural coordination patterns of phones used in their native language. But their recognition of these patterns is still broad and under-specified, at least for some phone categories, and does not reflect the linguistic

function of phonemic contrasts. At this point they appear to detect gestural properties in some non-native phones that are similar to the coordinated patterns they have begun to detect in native speech, but they are less able than adults to recognize the full pattern of similarities and discrepancies.

3) During the preschool years, the gestural coordination patterns of native phone categories become better differentiated, especially with respect to good versus less-good exemplars, but even by 4 years perception may not yet be fully organized at the level of phonemic contrast *per se*.

4) By adulthood, and probably much earlier, perception of speech sounds involves the recognition of linguistic structure at the level of phonemic contrasts, and unfamiliar sounds are "assimilated" to native phoneme categories on the basis of their articulatory-gestural similarities and discrepancies.

I have argued here that language-particular perceptual learning about speech involves the discovery of gestural coordination patterns that recur in the ambient language. It is this discovery that forms the basis of phonological development. This focus on the articulatory gestural properties perceived in native and non-native speech sounds should also extrapolate to the development of phonological organization in the child's speech productions. That is, as argued earlier, the ecological model of speech development posits that a common articulatory metric links perception and production. It follows that the child's emerging recognition of common gestural patterns in the ambient language should guide the development of language-specific phonological structure in his/her productions. In the final section of the chapter, we will examine this possibility in a case study of apparent phonological organization in a toddler's imitations of a diverse set of surface phonetic forms that realize a single phonemic contrast in American English.

Phonological behavior in early speech production

The study reported here found evidence of apparent phonological sophistication in a toddler's production of intervocalic alveolar stops at 20-22 months of age, based on phonetic and acoustic analyses of her imitations of a set of phonologically opaque adult targets. The American English adult targets were disyllables containing medial /d/ or /t/ followed by <-er>, <-le>, or <-en>. In these contexts, intervocalic alveolar stops typically appear in normal conversation as the restricted phonetic variants flap [ɾ], nasal release [d^{n}], or glottal stop [ʔ], (e.g., respectively, <wider> or <whiter>; <widen>; <whiten>). Although American English-speaking adults do not normally produce fully-released alveolar stops in these phonetic environments, the child consistently substituted full alveolar stops for the phonetic variants actually found in the adult targets. Nonetheless, the phonetic and acoustic characteristics of the child's responses differed among the diverse phonetic target forms, and distinguished between underlying /t/ and /d/. The analyses of the target utterances and of the child's imitations suggest that sensitivity to the articulatory properties of the target utterances, and/or articulatory constraints on her productions of the target phonemes in different phonetic environments, provided the basis for her behavior. In particular, the acoustically diverse allophones present in the adult targets nonetheless all involved alveolar contact coordinated with a release of obstruent constriction at some point in the vocal tract. I shall argue that the child's failure to exactly imitate the target utterances and the systematicity of her deviations reflect an important aspect of emerging phonological organization in her speech behavior. Specifically, the constraints provided by the articulatory information in the adult targets and those provided by the child's articulatory limitations (or preferences) determine how a phoneme would be realized in particular phonetic contexts and, consequently, how phonetically disparate forms may become related to a common underlying phonological category. This general line of reasoning is based on the model of articulatory phonology put forth by Browman & Goldstein (1986, 1989, in press), according to which phonological phenomena such as epenthesis, assimilation, and reduction can be understood simply as lawful consequences of the gestural organization of utterances.

This study was prompted by informal observations of my daughter Aurora at 20 months of age, when I noticed that she typically produced fully-released intervocalic alveolar stops while imitating adult words in which medial <t> was pronounced as a glottal stop [ʔ]. Because of the complex pattern of contextual effects that produce diverse phonetic realizations of phonemic <t> and <d> in intervocalic position in American English (see the preceding paragraph), I reasoned that Aurora's imitative responses to the differing phonetic forms of medial alveolar stops should reveal the characteristics of incipient phonological organization for those phonemes in her speech productions (Best, Goodell, & Wilkenfeld, in preparation).

Child phonology studies have typically excluded imitative responses from analysis, on the assumption that imitations would closely match the phonetic details of the adult target and thus would not reveal much about the intrinsic organization of the child's phonology (Leonard, Fey & Newhoff, 1981; Leonard, Schwartz, Folger & Wilcox, 1978). Yet Aurora's imitations were not exact phonetic replicas of the adult utterances; in fact, they differed *systematically* from the targets. Examining imitative responses offers several advantages over examining only spontaneous utterances: a) demands on memory and lexical access are minimized, b) unfamiliar and nonsense words can be presented to control for the influence of "phonological idioms" in familiar words (Moskowitz, 1980), (c) there is no doubt about the child's intended target, and (d) the properties of both the child's production and the immediate adult target can be directly compared. In addition, the standard approach in child phonology research has included only phonetic transcription of the child's utterances, and has not involved corollary acoustic analyses of the productions that might provide some converging evidence about their articulatory characteristics.

For these reasons, I recorded Aurora in three sessions between 20-22 months as she imitated familiar, unfamiliar, and nonsense target words containing intervocalic alveolar stops that were realized as phonetically diverse allophones. The targets were stress-initial disyllables containing intervocalic <t> or <d> produced as [ɾ], [ʔ] or [d^n]. Some words were familiar to the child and produced with familiar (American English: AE) or unfamiliar pronunciations (Cockney English: CE); others were unfamiliar words or nonsense words produced with AE or CE pronunciation. The medial stop was realized either as [ʔ] (e.g., AE <beaten> or CE pronunciation of <spittle>), or as [ɾ] (e.g., AE <batter>), or as the nasal release [d^n] (e.g., AE <widen>). The responses were elicited in the context of a vocal imitation game at home, during which Aurora was quite willing to provide citation-form repetitions (sometimes multiple tokens) of target words that had been presented in the sentence frame "Can you say ________?" We analyzed the phonetic and acoustic properties of both my targets and Aurora's responses for direct comparisons.

Both speakers' utterances were computer-digitized and the disyllables of interest were extracted into separate files. After discarding a small number of utterances that were not acoustically analyzable due to background noise (n_{child} = 10; n_{adult} = 4), broad phonetic transcriptions were made for the medial consonants in both the adult's (n = 51) and the child's (n = 52) remaining utterances. The transcriptions were conducted blind as to the context of the preceding and following utterances. These transcriptions were independently verified by a second listener. The majority of the utterances yielded identical transcriptions from the two transcribers for both speakers (child = 70%; adult = 100%). The greater difficulty in transcribing Aurora's utterances is typical of the generally decreased reliability for phonetic transcriptions of young children that has been noted in other early phonology studies. Discrepancies in the transcriptions for Aurora were resolved either by mutual agreement between the transcribers in a joint listening session (11% of her total utterances), or via tie-breaking by a third expert listener (19% of her utterances).

Several acoustic measurements were also taken from the intervocalic portion of each disyllable for both speakers. Pre-release silence (from end of first vowel to release burst) and voice onset time (VOT, measured as the time from release burst onset to beginning of glottal pulsing) were measured from the waveforms for tokens containing a stop release burst, or instead total duration of intervocalic silence was measured for tokens without a release burst. The latter measure provided a fairly objective index of the duration and timing of glottal devoicing. To index the timing of supralaryngeal (and/or glottal) closure gestures, we measured the total duration of closure, whether voiced or voiceless, for the intervocalic consonant of all utterances. Judgments of closure onset were based on a substantial, fairly abrupt decrease in signal amplitude and a qualitative change in voicing (according to changes in the waveform of the pitch pulses and/or to perceptual evidence of voice quality change) at the end of the first vowel, while judgments of closure offset were based on the onset of periodic voicing in the second vowel, rapid increase in amplitude and sometimes the presence of a release burst. To assess for any vowel-length differences associated with voicing differences in the medial stop, duration measurements of the first syllable were also taken.

According to phonetic transcriptions, the vast majority of Aurora's responses deviated from the phonetic properties of the adult targets. Instead, her responses were predominantly fully-released alveolar stops, corresponding to the phonological categories that underlie the diverse surface phonetic forms of the targets. Among Aurora's medial alveolar stops, the most frequent response to all three targets was /d/, which may reflect the relative ease of voiced alveolar stop production by children this age (Locke 1983; Kewley-Port & Preston, 1974) in addition to, or instead of, reflecting the phonological status of the adult targets. Nonetheless, the proportion of /d/ versus /t/ responses, and the pattern of less frequent responses, differed according to the phonological and phonetic (articulatory) properties of the targets. The voiced alveolar flap target [ɾ], which is the surface realization of medial /d/ or /t/ followed by syllabic [l̩] or [ɹ̩] in adult AE, yielded the largest proportion of /d/ responses (~ 80%), and no voiceless [t]s (the exceptions to /d/ were the voiced apical phones /ɾ/ and /ð/). In contrast, the glottal stop target [ʔ], which is the surface realization of /t/ followed by syllabic [n̩] in AE or CE and by syllabic [l̩] or [ɹ̩] in CE, elicited more /t/ responses (~ 35%) than any other target, which approached the proportion of /d/ responses for this target (~ 45%). The nasal (velar) release in target [dⁿ], which is the surface realization of an underlying /d/ before syllabic [ŋ̍], elicited an intermediate proportion of /d/s (~ 60%). The exceptions in the latter case included /t/, the velar stops /g/ and /k/, and the glottal /h/.

The acoustic measures also indicated that Aurora both deviated from the acoustic properties of the adult target utterances, and at the same time was systematically responsive to phonetic-articulatory differences among the surface forms of the targets. Note, first, that Aurora's alveolar stop responses almost always included stop release bursts, in contrast to the absence of bursts in nearly all of the adult targets—only two of the latter contained release bursts. More than half of the adult [ɾ] and [dⁿ] targets contained no intervocalic silence, and in those that did, the silent period was only 20-67 msec in duration. In contrast, the [ʔ] targets had consistent intervocalic silent periods of 50-150 msec. Consistent with the [ɾ]-[ʔ] asymmetry in the target pattern, the intervocalic silence in Aurora's /t/ and /d/ responses to [ʔ] targets were systematically longer than those to [ɾ] targets. This pattern is potentially consistent with either a gestural basis or an acoustic basis for Aurora's imitative responses. However, the intervocalic silent periods in her responses to the nasal release [dⁿ] targets were even longer than those to the glottal stops [ʔ], in direct opposition to the bias shown in the adult targets; thus, the child's responses in these cases cannot be a direct consequence of mimicking the acoustic properties of the targets. Instead, Aurora's responses to the [dⁿ] targets more likely reflect increased difficulty in her ability to execute the heterorganic alveolar and velar gestures of a medial [d] followed by a syllabic [ŋ̍].

Separate examination of Aurora's pre-release silent periods versus her post-release VOT measures offer some insight into the cause of the variations in her responses to the different targets. She produced the longest pre-release silent periods for the nasal targets (~ 100 ms), followed by glottal stops (~ 75 ms); her pre-release silences for flaps were substantially shorter (~ 20 ms). Thus, she appeared to use pre-release gap differences to systematically differentiate the three targets. However, she did not vary her post-release VOT systematically among the three targets. Her mean VOTs ranged from 40-50 ms for all three targets, and did not differ between the responses that were transcribed as /d/ vs. those transcribed as /t/ (see also Kewley-Port & Preston, 1974). Overall, these results indicate that the phonetic properties of the child's responses were achieved by means of different patterns of glottal timing than were the adult targets. The measurements suggest that she had greater control over the duration of the pre-release silent period than over the (post-release) VOT to instantiate intervocalic stop voicing distinctions.

We also examined evidence about the timing of closure itself, which is not adequately represented in the measures of silent intervals, especially for voiced stops. Given that children's utterances are longer than adults', the closure data were normalized by dividing the absolute closure duration by the length of the total utterance for both adult and child tokens. In all cases the child's absolute closure durations were longer than those of the adult. However, the child appeared to have produced systematic differences in ratios of closure/utterance-duration for the three target allophonic categories. These values were largest for responses to targets with medial glottal stops (~ .23), followed by those for the responses to nasal released targets (~ .18), and smallest for responses to flap targets (~ .12). The adult targets varied to a much lesser extent for absolute closure durations, and did not vary substantially in ratio of closure/utterance-duration. In the adult targets, the ratios for utterances containing nasal releases (~ .12) were nearly identical to those containing flaps and glottal stops (~ .11).

In American English and other languages, vowel length differences preceding an intervocalic or final stop often distinguish between voiced and voiceless versions of the stop. Therefore, differences in the length of the vowels preceding [ʔ] and [d^{n}] targets may have provided the child with additional information about their voicing difference. Moreover, these vowel length differences may also have been reflected in the child's productions. For these reasons, we measured vowel length in the first syllable of the adult and child utterances for all [ʔ]-[d^{n}] minimal pairs such as *batten* [bæʔn̩] and *badden* [bæːd̩^{n}]; the use of minimal pairs avoided confounding intrinsic vowel length differences with voicing-related vowel lengthening. Again, because children's utterances are longer than adults', the data were normalized by dividing the duration of the first vowel by the length of the total utterance. The adult targets showed vowel lengthening overall before /d/, but this pattern held only for the lax vowels /ʌ/, /ɪ/, and /æ/. Instead, the diphthongs /iɪ/ and /aɪ/ actually showed very slight shortening before /d/ relative to /t/. Aurora likewise showed overall vowel lengthening before /d/, but this held only for the diphthongs and not for the lax vowels, or exactly the opposite of the pattern found in the adult targets. Thus, again, the child appears to have achieved the voicing contrast in a different manner than was provided by the adult targets, rather than by simply mimicking the acoustic properties of the targets.

Overall, the results suggest some level of phonologically-relevant organization in Aurora's behavior, supported by the fact that systematic production patterns appeared across familiar and unfamiliar words and nonwords (some presented with unfamiliar CE pronunciation) that involved variations in phonetic contexts and in surface phonetic realizations of the underlying phonological categories. Because the study compared the phonetic transcriptions of the adult and child utterances with several acoustic measures that were intended to provide evidence about the articulatory gestures involved, the data may offer novel insights into the way in which young children begin to organize the articulatory/phonetic details of their productions in relation to abstract linguistic categories in the native phonological system. Specifically, the findings indicate:

1) that by the age of 20-22 months, Aurora had developed a systematic pattern of behavior that related the diverse surface phonetic forms of the medial consonants in the target utterances to underlying alveolar stops;

2) that Aurora's behavior pattern was more complex than a simple one-to-one mapping from a single allophone to a single phonological category, i.e., she did not simply imitate the simple acoustic properties of the targets. Instead, her behavior showed many-to-one mappings (she associated multiple surface forms to singular underlying phonemes) that nonetheless retained some articulatory/phonetic differentiation among the diverse, context-specific surface realizations of the categories;

3) and that the relationship between the allophones in the adult targets and Aurora's substitutions may be best understood in terms of articulatory characteristics of the targets and/or articulatory limitations on the child's productions, given the notable discrepancies between the acoustic properties of the targets and those of the child's imitations.

The complexity of these patterns in Aurora's productions is demonstrated not only by her nearly-consistent substitution of alveolar stops for glottal stops, nasal releases, and flaps, but also by the differences in intervocalic silence and in closure intervals that she maintained among the stops she substituted for [ɾ], for [ʔ] and for [d^{n}].

The characteristics of the child's substitutions indicate a behavior pattern that could be considered to reflect the beginnings of phonemic organization, but an organization that is still immature and quite different from that seen in adult speech behavior. The results suggest that this behavior pattern is based on articulatory properties of the phones and phonological categories investigated. Consider that all of the intervocalic stop targets involve forward movement of the tongue tip into alveolar contact or approximation, either because of an intrinsic alveolar gesture in the target segment itself ([ɾ], [d^{n}], [ʔ]) and/or because of alveolar gestures in the following syllable ([n̩], [l̩], or [ɹ̩]). Accordingly, the vast majority even of Aurora's substitutions were consonants involving apical contact (including /n/ in response to [ʔ], and the linguo-dental /ð]/ in response to [ɾ]). She could have done otherwise—she was producing a broad array of stops, nasals, fricatives and approximants at various places of articulation at this age, including glottal stops (e.g., <uh-oh> as [ʌʔo^{ʊ}]). In addition, other articulatory properties of the targets could be related to gestures other than alveolar stop maneuvers. Two of the targets ([ʔ], [d^{n}]) incorporate posterior vocal tract gestures (glottal stop, velar constriction) in addition to alveolar articulations, which could appropriately have been substituted by posterior consonants. In fact, of the "other" responses, those to [ʔ] and [d^{n}] were all posterior gestures (/g/, /k/, /h/), while the only "other" response to [ɾ], which did not involve a posterior gestural component, was the linguo-dental (/ð/). Nonetheless, it is important to remember that the vast majority of Aurora's responses were indeed fully released alveolar stops.

It thus appears that Aurora was sensitive to the gestural properties of the target words, even if she did not mimic them precisely. Specifically, she appears to have been sensitive to the main place(s) of constriction in the vocal tract, and to characteristics of the associated glottal gestures. Furthermore, she was able to incorporate information about those properties into her productions, so that she could both relate her productions to linguistic categories in her native language, and arrive at a phonetic realization for them within her articulatory limitations. Her primary difficulty was apparently a failure to incorporate into her imitations the precise temporal coordination among the supralaryngeal and the glottal gestures—the interrelations among discrete gestures of different articulators—that was provided in the adult targets. These findings are compatible with the ecological approach to speech development discussed earlier in the chapter, and with the principles of Browman and Goldstein's model of articulatory phonology (1986, 1989, in press).

These imitation data carry important implications about the development of phonological organization in children's speech productions, which appear to complement our findings on perceptual assimilation of non-native phones. That is, the pattern of systematic, context-specific articulatory variations in Aurora's productions, along with their relation to the voicing categories of alveolar stops, suggest that she related diverse allophonic realizations to common phonological categories. Her phonological categories, however, were organized differently than in adult speech productions, and appear to be under-differentiated with respect to the variations in gestural coordination that are found among allophones in the native phonology. This pattern would seem most consistent with the category recognition hypothesis discussed earlier regarding developmental reorganization in infants' perception of non-native contrasts.

CONCLUDING REMARKS

This chapter has examined the way in which the infant may come to relate the phonetic details of speech in his/her language environment to the more abstract linguistic categories of the native phonological system. To this end, I have presented a model of how the infant may move from perceiving general information about phonetic contrasts in speech during the first six months of life to discovering language-particular patterns that ultimately correspond to the phonemes and phonemic contrasts that guide speech perception and production in the mature language-user. This discovery, in turn, influences the properties perceived in non-native speech sounds. I have also presented complementary information about the possible relations between the phonetic properties of adult utterances in the native language and the emergence of phonologically-relevant organization in a young language-learning child's speech productions.

The line of reasoning developed in this chapter is compatible with the premise that the recognition of phonemes as specifically linguistic elements, which convey meaningful contrasts and are functionally organized within a phonological system, develops only gradually as the child builds a lexicon (see Studdert-Kennedy, 1987, 1989; Flege, 1990; cf. Jusczyk, 1986, in press). According to that view, phonemic segments are differentiated from words, rather than being pre-existing elements that are concatenated into words, and the child's phonological system emerges in accord with the principles of self-organizing systems (e.g., Studdert-Kennedy, 1987; 1991).

I have suggested that the means by which this phonemic development takes place in perception and production is through the young speaker-hearer's detection of information in speech about the articulatory events that produced the signal. This ecological view of speech development, which was based upon James Gibson's general ecological theory about perceptual systems, thus posits a common articulatory link between perception and production of speech. That commonality should provide obvious benefits for the young child's acquisition of a native language. The data summarized here support the proposed model of phonological development, and are compatible with the

ecological theoretical perspective described. I would argue that this ecological perspective on speech offers important and unique insights about the relation between perception and production in the development of spoken language. Future research must resolve the full time-course of development of mature phonological organization in speech perception and production.

REFERENCES

Aitchinson, J. (1983) *The articulate mammal.* New York: Universe Press.

Aslin, R. N., & Pisoni, D. B. (1980a). Some developmental processes in speech perception. In G. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology, Volume 1: Production* New York: Academic Press.

Aslin, R. N., & Pisoni, D. B. (1980b). Effects of early linguistic experience on speech discrimination by infants: A critique of Eilers, Gavin, & Wilson (1979). *Child Development, 51,* 107-112.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development, 52,* 1135-1145.

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In P. H. Mussen (Series Ed.), M. M. Haith & J. J. Campos (Vol. Eds.), *Handbook of child psychology: Infancy and developmental psychobiology* (Vol. 2, pp. 573-687). New York: Wiley.

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior and Development, 10,* 387-416.

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development, 59,* 197-209.

Best, C. T. (1990). Adult perception of nonnative contrasts differing in assimilation to native phonological categories. *Journal of the Acoustical Society of America, 88,* S177.

Best, C. T. (1984). Discovering messages in the medium: Speech and the prelinguistic infant. In H. E. Fitzgerald, B. Lester, & M. Yogman (Eds.), *Advances in pediatric psychology* (Vol. 2, pp. 97-145). New York: Plenum.

Best, C. T. (1991a). Language-specific effects on perception of discourse prosody categories by 2-4 month olds. Presented at the International Conference on Event Perception and Action, Amsterdam, The Netherlands, August.

Best, C. T. (1991b). Phonetic influences on the perception of nonnative speech contrasts by 6-8 and 10-12 month-olds. Presented at the meeting of the Society for Research in Child Development, Seattle WA, April.

Best, C. T., Goodell, E., & Wilkenfeld, D. (in preparation). *Phonologically-motivated substitutions in a 20-22 month old's imitations of intervocalic alveolar stops.*

Best, C. T., Levitt, A., & McRoberts, G. W. (1991). Examination of language-specific influences in infants' discrimination of prosodic categories. *Proceedings of the XIIth International Congress of Phonetic Sciences, Vol. 4* (pp. 162-165) Aix-en Provence: Universite de Provence, France.

Best, C. T., & McRoberts, G. (1989, April). Phonological influences on infants' perception of two non-native speech contrasts. Paper presented at the *Society for Research in Child Development,* Kansas City, MO.

Best, C. T., McRoberts, G. W., Goodell, E., Womer, J. S., Insabella, G., Kim, P., Klatt, L., Luke, S., & Silver, J. (1990). Infant and adult perception of nonnative speech contrasts differing in relation to the listener's native phonology. *Infant Behavior and Development.*

Best, C. T. McRoberts, G. W., & Sithole, N. N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345-360.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics, 29,* 191-211.

Brière, E. (1966). An investigation of phonological interference. *Language, 42,* 769-796.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3,* 219-252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 62.*

Browman, C. P., & Goldstein, L. (in press). Articulatory phonology: An overview. *Phonetica.*

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics, 7,* 207-240.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961-970.

Dent, C. (1989). An ecological approach to language development: An alternative functionalism. *Developmental Psychobiology, 22.*

Diehl, R., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology.*

Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *Journal of the Acoustical Society of America, 61,* 1321-1336.

Eilers, R. E., Gavin W. J., & Oller, D. K. (1982). Cross linguistic perception in infancy: The role of linguistic experience. *Journal of Child Language, 9,* 289-302.

Eilers, R. E., Gavin W. J., & Wilson, W. R. (1979). Linguistic experience and phonemic perception in infancy: A cross-linguistic study. *Child Development, 50,* 14-18.

Eilers, R. E., Gavin W. J., & Wilson, W. R. (1980). Effects of early linguistic experience on speech discrimination by infants: A reply. *Child Development, 51,* 113-117.

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research, 18,* 158-167.

Eilers, R. E., Morse, P. A., Gavin, W. J., & Oller, D. K. (1981). The perception of voice-onset-time in infancy. *Journal of the Acoustical Society of America, 70,* 955-965.

Eilers, R. E., & Oller, D. K. (1988). Precursors to speech: What is innate and what is acquired? *Annals of Child Development, 5,* 1-32.

Eilers, R. E., Oller, D. K., Bull, D. H., & Gavin, W. J. (1984). Linguistic experience and infant speech perception: a reply to Jusczyk, Shea and Aslin (1984). *Journal of Child Language, 11,* 467-475.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research, 20,* 766-780.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1979). Speech discrimination in the language-innocent and the language-wise: A study in the perception of voice onset time. *Journal of Child Language, 6,* 1-18.

Eimas, P. D. (1974). Linguistic processing of speech by infants. In R. L. Schiefelbusch & L. L. Lloyd (Eds.), *Language perspectives—Acquisition, retardation and intervention.* Baltimore: University Park Press.

Eimas, P. D. (1975). Speech perception in early infancy. in L. B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition.* New York: Academic Press.

Eimas, P. D. (1978). Developmental aspects of speech perception. in R. Held, H. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology: Vol. VIII. Perception*. New York: Springer-Verlag.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303-306.

Fant, (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. In J. S. Perkell & D. H. Klatt (Eds.) *Invariance and variability of speech processes* (pp. 36-53). Hillsdale, NJ: Lawrence Erlbaum Associates.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America, 76,* 692-707.

Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Psycholinguistics, 8,* 162-177.

Flege, J. E. (1988). The production and perception of speech sounds in a foreign language. In H. Winitz (Ed.), *Human communication and its disorders: A review 1988.* Norwood, NJ: Ablex.

Flege, J. E. (1990). Perception and production: The relevance of phonetic input to L2 phonological learning. In C. Ferguson & T. Huebner (Eds.) *Crosscurrents in second language acquisition and linguistic theories.* Philadelphia: John Benjamins.

Flege, J. E. (in press). Speech learning in a second language. In C. Ferguson, L. Menn, & C. Stoel-Gamon (Eds.), *Phonological development: Models, research, and applications.* Parkton, MD: York Press.

Flege, J. E, & Eefting, W. (1986). Linguistic and developmental effects on stop perception and production by native speakers of English and Spanish. *Phonetica, 43,* 323-347.

Flege, J. E., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics, 15,* 67-83.

Flege, J. E., McCutcheon, M., & Smith, S. (1987). The development of skill in producing word-final English stops. *Journal of the Acoustical Society of America, 82,* 433-447.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3-28.

Fowler, C. A. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology, 1,* 145-160.

Fowler C. A. (1991). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America, 88,* 1236-1249.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics, 48,* 559-570.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 816-828.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123-139). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, C. A., & Turvey, M. T. (1978). Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. Stelmach (Ed.) *Information processing in motor control and learning.* New York: Academic Press.

Gardner, B. T., & Gardner, R. A. (1973). Evidence for sentence constituents in the early utterances of child and chimpanzee. *Journal of Experimental Psychology: General, 104,* 244-267.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston, MA: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghton-Mifflin.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32-41.

Gillette, S. (1980). Contextual variation in the perception of L and R by Japanese and Korean speakers. *Minnesota Papers in Linguistics and the Philosophy of Language, 6,* 59-72.

Goodell, E., & Studdert-Kennedy, M. (1990). From phonemes to words or words to phonemes: How do children learn to talk? Paper presented at the *International Conference on Infant Studies,* Montreal, April 19-22.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia, 9,* 317-323.

Gottlieb, G. (1981). The roles of early experience in species-specific perceptual development. In R. N. Aslin, J. R. Alberts, & M. R. Peterson (Eds.), *Development of perception: Psychobiological perspectives, Vol. 1, Audition, somatic perception, and the chemical senses.* New York: Academic Press.

Grieser, D. A., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25,* 577-588.

Insabella, G. (1990). Four-year-olds' perception of non-native phonetic contrasts. Unpublished senior honors thesis, Wesleyan University, Middletown CT.

Insabella, G., & Best, C. T. (1990). Four-year-olds' perception of nonnative contrasts differing in phonological assimilation. Paper presented at meeting of the *Acoustical Society of America,* San Diego, November 26-30.

Jusczyk, P. W. (1981). Infant speech perception: A critical appraisal. In P. D. Eimas & J. A. Miller (Eds.) *Perspectives in the study of speech.* Hillsdale NJ: Lawrence Erlbaum Associates.

Jusczyk, P. W. (1985). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion* (pp. 199-229). Hillsdale, NJ: Lawrence Erlbaum Associates, .

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Jusczyk, P. W. (in press). Speech perception and the lexicon. In H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* Cambridge, MA: MIT Press.

Jusczyk, P. W., Shea, S. L., & Aslin, R. N. (1984). Linguistic experience and infant speech perception: A re-examination of Eilers, Gavin and Oller (1982). *Journal of Child Language, 11,* 453-466.

Kewley-Port, D., & Preston, M. S. (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics, 2,* 195-210.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2, Perception* (pp. 41-66). New York: Academic Press.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception,* (Vol. 2, pp. 275-382). New York: Academic Press.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93-107.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218,* 1138-1144.

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development, 7,* 361-381.

Ladefoged, P. (1981). *A course in phonetics* (2nd ed.). New York: Harcourt, Brace & Jovanovich.

Lalonde, C., & Werker, J. (1990). Cognitive/perceptual integration of three skills at 9 months. Paper presented at the *International Conference on Infant Studies,* Montreal, April 19-22.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology, 20,* 215-225.

Leonard, L. B., Fey, M. E., & Newhoff, M. (1981). Phonological considerations in children's early imitative and spontaneous speech. *Journal of Psycholinguistic Research, 10,* 123-133.

Leonard, L. B., Schwartz, R. G., Folger, M. K., & Wilcox, M. J. (1978). Some aspects of child phonology in imitative and spontaneous speech. *Journal of Child Language, 5,* 403-415.

Levitt, A., Jusczyk, P., Murray, J., & Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 361-368.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 245,* 489-494.

Lieberman, P. (1975). *On the origins of language: An introduction to the evolution of human speech.* New York: MacMillan Publishing Co.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Universals Workshop.* The Hague: Mouton.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments on comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences.* Prague: Academia.

Locke, J. L. (1983). *Phonological acquisition and change.* New York: Academic Press.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24,* 253-257.

MacKain, K. S. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child Language, 9,* 527-542.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2,* 369-390.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. *Science, 219,* 1347-1349.

Macken, M.A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua, 49,* 11-49.

Macken, M. A., & Ferguson, C. A. (1983). Cognitive aspects of phonological development: Model, evidence, and issues. In K. E. Nelson (Ed.), *Children's language.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Mann, V. A. (1980). Influence of of preceding liquid on stop consonant perception. *Perception & Psychophysics, 28,* 407-412.

Mann, V. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of "l" and "r." *Cognition, 24,* 169-196.

Massaro, D. M. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Mattingly, I. G., & Liberman, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function.* New York: Academic Press.

McClasky, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Effects of transfer of training on identification of a new linguistic contrast in voicing. *Research on Speech Perception Progress Report (Indiana University), 6,* 205-234.

McCune, L., & Vihman, M. (1987). Vocal motor schemes. *Papers and Reports in Child Language Development, 26.*

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N. Bertoncini, J., & Amiel-Tison, C. A. (1988). A precursor of language acquisition in young infants. *Cognition, 29,* 143-178.

Menn, L. (1971). Phonotactic rules in beginning speech. *Lingua, 26,* 225-251.

Menn, L. (1978). Phonological units in beginning speech. In A. Bell & J. Hooper (Eds.), *Syllables and segments.* Amsterdam: North Holland.

Menn, L. (1986). Language acquisition, aphasia and phonotactic universals. In F. R. Eckman, E. A. Moravcsik, & J. R. Wirth (Eds.), *Markedness* (pp. 241-255). New York: Plenum Press.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18,* 331-340.

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics, 9,* 283-303.

Moskowitz, B. A. (1980). Idioms in phonology acquisition and phonological change. *Journal of Phonetics, 8,* 69-83.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32(1),* 120-132.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology, Volume 1: Production.* New York: Academic Press.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research, 5,* 261-283.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 297-314.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55,* 328-333.

Pisoni, D. B., Logan, J. S., & Lively, S. E. (in press). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* Cambridge, MA: MIT Press.

Premack, D. (1971). Language in chimpanzee? *Science, 172,* 808-822.

Rosenblum, L. (1987). Towards an ecological alternative to the motor theory of speech perception. *Perceiving-Acting Workshop Review (Tech. Rep. CESPA, University of Connecticut), 2,* 25-28.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3,* 243-261.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 1: Production.* New York: Academic Press.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36,* 131-145.

Streeter, L. A. (1976a). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature, 259,* 39-41.

Streeter, L. A. (1976b). Kikuyu labial and apical stop discrimination. *Journal of Phonetics, 4,* 43-49.

Studdert-Kennedy, M. (1981). The emergence of phonetic structure. *Cognition, 10,* 301-306.

Studdert-Kennedy, M. (1985). Perceiving phonetic events. In W. H. Warren & R. E. Shaw (Eds.), *Persistence and change.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Studdert-Kennedy, M. (1986a). Development of the speech perceptuomotor system. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech.* New York: Stockton Press.

Studdert-Kennedy, M. (1986b). Sources of variability in early speech development. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 67-84). London: Academic Press.

Studdert-Kennedy, M. (1989). The early development of phonological form. In C. von Euler, H. Forssberg & H. Lagercrantz (Eds.), *Neurobiology of early infant behavior* (pp. 287-301).Basingstoke, England: MacMillan.

Studdert-Kennedy, M. (1991). Language development from an evolutionary perspective. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Language acquisition: Biological and behavioral determinants.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Tahta, S., Wood, M., & Loewenthal, K. (1981). Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech, 24,* 265-272.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38,* 579-590.

Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence? *Science, 206,* 809-902.

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology, 9,* 91-96.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by adults and infants. *Child Development, 47,* 466-472.

Waterson, N. (1971). Child phonology: A prosodic view. *Journal of Linguistics, 7,* 179-211.

Werker, J. F. (1989). Becoming a native listener. *American Scientist, 77,* 54-59.

Werker, J. F. (in press). Cross-language speech perception: Developmental change does not involve loss. In H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* Cambridge, MA: MIT Press.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52,* 349-355.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672-683.

Werker, J., & Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37,* 35-44.

Werker, J. F., & Tees, R. C. (1984a). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866-1878.

Werker, J. F., & Tees, R. C. (1984b). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

FOOTNOTES

[*] Chapter to appear in H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* Cambridge, MA: MIT Press (in press).

[†] Also Wesleyan University.

[1] This is not meant to imply that infants initially have universal abilities to perceive all contrasts from all languages, although this has sometimes been claimed or implied. Certain contrasts, both native and nonnative, may be more difficult than others for young infants to discriminate (e.g., Aslin, Pisoni, Hennessy, & Perey, 1981; Eilers & Minifie, 1975; Eilers et al., 1981; Eilers, Wilson, & Moore, 1979; Kuhl, 1980) [see further discussion in section on *Language-particular developments in infant speech perception*]. The point here is only that initially, infants' discrimination of segmental contrasts does not yet appear to be constrained by the particular language environment.

[2] One alternative is that the auditory system begins and remains physiologically incapable of registering certain acoustic properties of speech unless the listeners' environment provides exposure to those properties, which would then induce sensitivity (presumably during some critical developmental period). This would be a strictly sensory-neural version of the Induction Hypothesis formulated by Aslin and Pisoni (1980a; see also Pisoni, Logan, & Lively, in press) to explain one possible form of experiential effect on perceptual development, which was based on Gottlieb's (1981) model of visual and auditory development in ducklings. Another, alternative effect of auditory experience has been suggested by Eilers and colleagues (e.g., Eilers & Oller, 1988; Eilers, Wilson & Moore, 1979)—that some phonetic contrasts are weak in acoustic salience and hence initially difficult for the infant to discriminate, but that (relatively frequent) exposure to these can improve discrimination. The logical problem with the absolute notion of sensory-neural induction is that if the sensory system is *incapable* of registering an acoustic property, the exposure is intrinsically ineffectual and incapable of inducing sensitivity. To illustrate, the human visual system cannot register ultraviolet wavelengths as visible light, and our abundant exposure to ultraviolet light from both natural and artificial sources never induces visual sensitivity to those wavelengths. The acquisition of *perceptual* ability to discover previously unrecognized organization in stimulation is another matter, however, so long as the sensory system is already capable of registering the supporting physical evidence. The emergence of such perceptual abilities certainly does occur, associated with adjustments of selective attention (Gibson, 1966; Gibson & Gibson, 1955). In these cases, the term "perceptual learning" is preferable to "induction." Sensory capacity is necessary though not sufficient for perceptual learning; but perceptual learning cannot induce sensory capacity. In other words, if the system *cannot* register a stimulus property, experience will not change that fact. If the system can register the property, but the perceiver does not initially *recognize* the pattern of information it conveys, then experience can lead to perceptual learning. Confusion about induction may arise within a theoretical model to the extent that it

conflates experiential effects on physiological mechanisms with experiential effects on perceptual skills (i.e., selective attention).

[3] Space limitations prevent a full recapitulation of the discussion here; for additional details, the reader is referred to the original papers.

[4] A very similar question about adult perception and production of non-native sounds has been addressed in the work of Flege (e.g., 1988, in press; Flege & Eefting, 1986, 1987), whose discussion of the issue is compatible in some respects with the view presented here, but divergent in other respects.

[5] The diversity of possible assimilations is due to the fact that Zulu /ɬ/ shares partial articulatory commonalities with each of these (and perhaps other) English consonants, and is simultaneously discrepant from each of them on other articulatory properties, i.e., the gestural *coordination* is unfamiliar with respect to English. As stated earlier, listeners should show sensitivity to both similarities and discrepancies between the non-native phone and native categories. Furthermore, they may vary regarding which similarities/discrepancies capture their attention, hence they may differ regarding which of several possible native categories assimilates a given nonnative phone.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 31-40

Intonational Differences Between the Reduplicative Babbling of French- and English-learning Infants*

D. H. Whalen, Andrea G. Levitt,† and Qi Wang††

The two- and three-syllable reduplicative babbling of five French-learning and five English-learning infants (0;5 to 1;1) was examined in two ways for intonational differences. The first measure was a categorization into one of five categories (Rising, Falling, Rise/Fall, Fall/Rise, Level) by expert listeners. The second was the fundamental frequency (F0) from the early, middle and late portion of each syllable. Both measures showed significant differences between the two language groups. 65% of the utterances from both groups were classified as either Rising or Falling. For the French children, these were divided equally into the rising and the falling categories, while 75% of those utterances for the English children were judged to have falling intonation. Proportions of the other three categories were not significantly different by language environment. In both languages, though, three-syllable utterances were more likely to have a complex contour than two-syllable ones. Analysis of the F0 patterns confirmed the perceptual assessment. Several aspects of the target languages help explain these intonational differences in prelinguistic babbling.

INTRODUCTION

The transition from babbling to language has been the focus of much recent research (e.g., Thevenin, Eilers, Oller, & Lavoie, 1985; Vihman, Ferguson & Elbert 1986; Vihman, Macken, Miller, Simons & Miller, 1985). This wide range of studies has an equally wide range of implications. On the one hand, there is substantial evidence of universal developments, which seem to reflect a maturational process independent of the language environment (e.g., Locke, 1983). On the other hand, there is equally compelling evidence for language-specific changes, where the character of the babble begins to reflect the composition of the adults' language (e.g., de Boysson-Bardies, Sagart & Durand, 1984). Given the current evidence for the continuity between prelinguistic utterances and the acquisition of the ambient language, it could hardly be otherwise than that there would be evidence of both kinds of processes. The task now is to further trace the emergence of various attributes of the adult language.

Previous studies of both maturational and language-specific influences in the babbling of infants from different linguistic communities have rarely combined both perceptual and instrumental analyses. These two approaches to the study of prelinguistic speech complement one another well. Perceptual studies with an instrumental check are useful in determining the physical properties of the stimulus that are the source of the perceptual judgment. Furthermore, instrumental analyses have sometimes revealed linguistically-predicted differences that were not captured in the original transcriptions (Macken & Barton, 1980).

One aspect of babbling that deserves special attention is its intonation. Infants often appear to imitate intonation patterns before they imitate segmental structure (Crystal, 1973; Lewis, 1936; Weir, 1966). There is some evidence that children can begin mastering the F0 changes that represent lexical tone before they command the segmental distinctions of their language environment (Li & Thompson, 1976). French children use intonation to express different meanings at the one word stage (Marcos, 1987). Despite this early suc-

---

The research reported here was supported by NIH grant DC-00403 to Haskins Laboratories and Catherine T. Best. We thank Catherine T. Best, Carol A. Fowler, Janet F. Werker and two anonymous reviewers for helpful comments.

cess with intonation, control of syntactically appropriate intonation is a later development (Crystal, 1979; Weir, 1966). Furthermore, heightened F0 excursion is a salient characteristic of motherese, a speaking style that infants pay particular attention to (Ferguson, 1964; Fernald, Taeschner, Dunn, Papoušek, de Boysson-Bardies & Fukui, 1989; Garnica, 1977). It has been suggested that it is indeed the intonation pattern of the F0 that attracts infants' attention (Fernald & Kuhl, 1987).

Do the F0 contours in the child's babbling reflect the predominant patterns in the adult language? Very little work to date has been addressed to this question. The present work will make a start by examining the reduplicative babbling of five children from French environments and five from English environments. These languages differ not only in their inventories of contours (e.g., Cruttenden, 1986; Pierrehumbert, 1980; for English; Kenning, 1979; Leach, 1988 for French) but also in which of their intonation contours are most typical. Infants will be more likely to incorporate those patterns that they hear more frequently, leading us to examine more statistically-oriented studies. Delattre (1961), classifying patterns from one French and one American female speaker into rising or falling, found 90% of American English contours were falling, while 87% of French contours were rising. Pike (1945:155), using five categories, finds a preponderance (53.7%) of falling contours at all sentence positions for American English, but very few rising contours (9.6%). A further salient difference between the two languages is that the contour signalling the speaker's intent to continue is usually rising in French but falling, or falling then rising, in English (Delattre, 1961; Rossi, 1980; Grover, Jamieson, & Dobrovolsky, 1987). This gives a different sound to the two languages, despite the presence of a final declarative fall for both of them.

Our first goal, then, was to determine whether intonation differences, if they exist between these two groups of infants, mirror the marked differences between falling and rising F0 in the two language environments. This was assessed both perceptually, by having expert listeners classify the intonation patterns, and instrumentally, by an analysis of the actual F0. A comparison of the two measurements is made possible by our restriction to reduplicative babbles, given that infants as well as adults have F0s which depend in part on the quality of the vowel they are producing (Bauer, 1988). Such intrinsic F0 effects make the comparison of the two measurements difficult if there are different vowels in the utterances, since the different vowels would, in some cases, have different F0s and yet sound as if they had the same pitch (Lehiste, 1970; Silverman, 1986). The acoustic analysis by itself is insufficient for the purpose of delimiting perceptually relevant changes in intonation (Crompton, 1980:152; Leach, 1988:125).

METHOD

Subjects. The subjects for the present study were natives of France and the United States, growing up in monolingual French or English environments respectively. The average age of the children in the first recording used was 0;7.10 and the last was at an average of 0;11.3 (ranging from 0;5 to 0;13: see Table 1).

Stimuli. The stimuli for the present study were selected from recordings made at weekly intervals by the parents of the children. Each infant was recorded in the home on a cassette tape recorder using a high quality microphone. Individual recording sessions lasted approximately 10-20 minutes. The parents were to choose a time when the child was likely to be alert and happy. To the extent possible, the microphone was held 20 cm from the baby. If necessary, the parent could attempt to induce babbling by interacting with the child (stopping, of course, when the infant began). Additional comments about the session were recorded by the parent on a form provided with each tape.

Table 1. *Description of the source of the 156 stimuli.*

	English-learning infants					French-learning infants				
initials:	AB	MM	VB	CR	MA	EC	MS	MB	NM-B	JZ
sex:	F	M	F	M	M	M	M	M	F	M
no. 2-sylls	11	10	13	3	11	24	10	11	5	0
no. 3-sylls	4	13	4	3	4	11	8	7	1	3
first age	0;7	0;7	0;7	0;9	0;8	0;6	0;7	0;8	0;9	0;5
last age	0;11	1;0	0;11	0;10	1;0	1;0	1;0	0;11	1;1	0;7

All the usable two- and three-syllable reduplicative babbles were selected for this study. This was done by extracting them from computer digitizations. The recordings were low-pass filtered at 4.8 kHz and sampled at 10 kHz via the Haskins PCM system (Whalen, Wiley, Rubin, & Cooper, 1990). While all the utterances in the sessions have been transcribed for other purposes, we chose only the reduplicative babbles for this study. As mentioned in the introduction, we chose to eliminate vowel changes as a source of F0 confound by keeping vowels constant within an utterance. We limited the length of the utterance to three syllables both because longer utterances can have more complicated intonations than could be easily compared and because the number of such utterances is too small for systematic comparisons. (At least 750 ms of silence occurred after the last syllable.) With these criteria, we obtained 156 stimuli. These were almost equally divided between the two language groups (76 for the English children, 80 for the French), though the differences among the individuals in the number of usable utterances were large (see Table 1).

Listeners. The listeners were the first two authors and two colleagues familiar with infant vocalizations but not with the stimuli at hand. Since the third author was in charge of selecting the stimuli, the authors were able to make their judgments on the stimuli blind to the language background of the infant. All the listeners were native speakers of a variety of American English.

Procedure. The two- and three-syllable utterances were recorded on audio tape for presentation to the listeners. Each was repeated three times with 1 sec between repetitions. After the third repetition, there was a 4.5 sec pause for the listener to write down a response. Two randomizations, each containing one triplet of each of the 156 stimuli, were prepared. The listeners heard the two tapes in balanced order.

Each intonation pattern was to be classified as Rising, Falling, Rise/Fall, Fall/Rise, or Level. It was made clear that the contour did not have to be within one syllable. Thus two syllables which had a level pitch within-syllable but with a difference between syllables should be judged to have a contour. However, a sizable contour on a single syllable could also be considered in the overall judgment. Similarly, it was not necessary for there to be no pitch change at all for a judgment of "flat." Rather, those utterances which lacked large pitch changes and had no clear pattern to the small ones that were present were also to be classified as "flat." Even in uncertain cases, a single response was to be written. For some of the complex patterns, no single pattern could be expected to be the majority decision.

Results

There was good agreement on the intonation of most of the stimuli. In 60% of the cases, one category accounted for at least 75% of the judgments from all the listeners, and in 91% there was one category that accounted for at least 50% (the "majority" decision).

As can be seen in Figure 1, there were differences in the judged intonation of the babbles from the children in the two different language environments. For the French children, most of the utterances had simple rising or falling patterns, and about equal numbers of each. For the English children, the falling pattern was a sizable plurality. For both groups, the other three categories accounted for about a third of the responses, with no clear pattern within or across languages. The results were almost identical for the two-syllable and three-syllable utterances. The only noticeable shift was toward slightly more rise/fall for the three syllable, and more flat for the two-syllable utterances.

The eight judgments on each stimulus were submitted to a repeated measures analysis of variance. To code the contours numerically, we counted the flat contour as 0 and deviations from that in two directions. Thus the rise/fall was coded as 1, and the rise as 2, while the fall/rise was coded as -1 and the fall as -2. The four listeners and their two judgments per stimulus were treated as within factors for the 156 utterances, which were grouped by the between factors of Language environment and Number of syllables.

Language environment, confirming the obvious patterns in Figure 1, was a significant factor ($F(1,152) = 5.53, p < .05$). The number of syllables did not affect the intonation judgments ($F(1,152) < 1.0$, n.s.). In the coding scheme, the difference between flat and rise/fall (the only difference noted in the percentages of judgments) is a small one and so does not contribute much to the variance accounted for by this factor. The interaction of these two factors was not significant.

Despite the sizable agreement among listeners, there were still significant differences among them ($F(3,456) = 2.88, p < .05$), and some differences for the listeners' judgments on the two languages ($F(3,456) = 3.58, p < .05$).

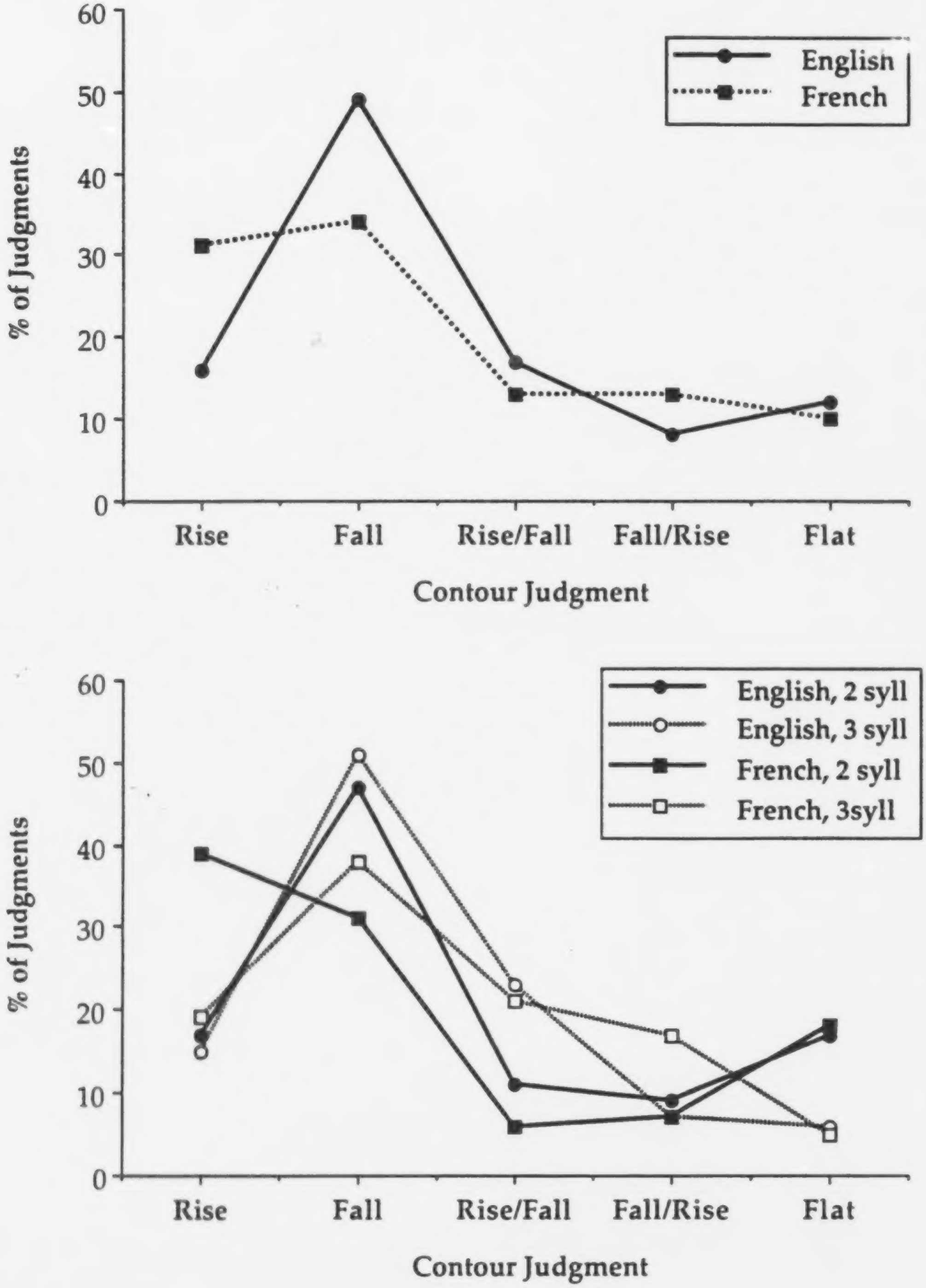


Figure 1. **Percentage of listener judgments for each of the five intonation contours. Each function represents the total for the utterances from the five children learning each language (French and English).**

Although all four listeners had more falling (i.e., negative) responses for the English than for the French, they differed in the degree of disparity between the two languages. No other interactions with this factor were significant, nor were any involving the difference between the two repetitions.

ACOUSTIC ANALYSIS

Data Reduction and Analysis. The focus of our acoustic analysis was fundamental frequency, which, though not the only determinant of prosody, is the major one (Cruttenden, 1986:2-3). F0 values for each syllable of the 156 stimuli were obtained for comparison with the perceptual judgments. For the analysis, the medians of the values for each third of the syllable were obtained. While such a selection has the drawback of ignoring duration differences between syllables, it has the advantage that there are the same number of measurements for each syllable. Each of the utterances was analyzed with the API command of the Interactive Laboratory System package. Once the values were obtained, they were compared with an auditory judgment of the closeness of fit. In many cases, the value reported is actually one half or one third the true value. In these cases, the numbers obtained were simply multiplied by an appropriate amount. In some cases, no pitch was detected where there clearly was some. In these cases, an average measurement was obtained by hand-delimiting ten pitch periods and finding the average F0. Such changes and replacements were done for 30% of the syllables in the corpus.

ACOUSTIC RESULTS

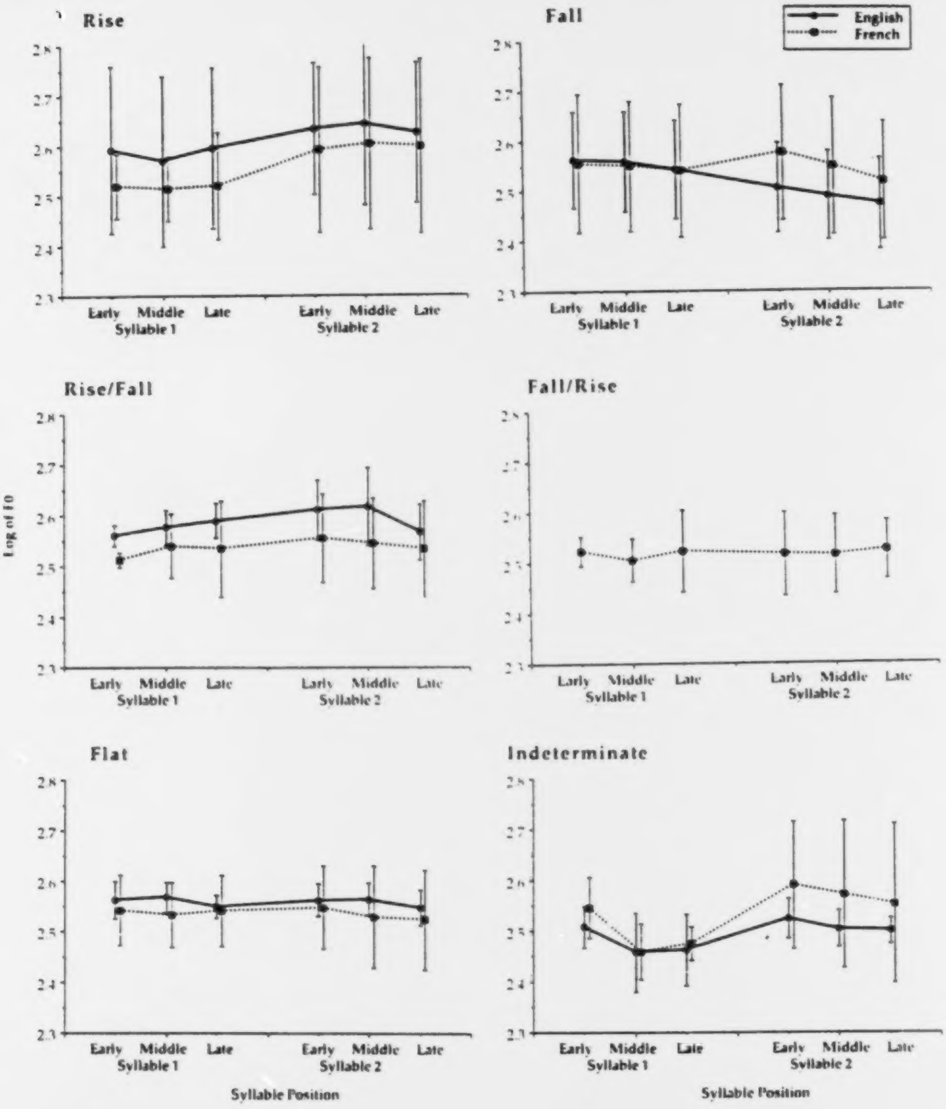
Before computing intonation contours based solely on the F0 pattern, we determined the fit between the listener judgments and the computed F0 pattern. For this analysis, the utterances were divided according to the category that received at least half of the listener judgments. 91% of the utterances had such a majority judgment, and the other 9% were classified as "Indeterminate." Pseudo-contours were obtained by averaging the mean of the logs of the three F0 values per syllable. Logs of the values in Hz were used to make the changes in F0 more equivalent across the range of F0s. Other researchers have used semitones to analyze F0 ranges. Since the semitone conversion requires a baseline, we used the more independent log scale. The semitone scale is logarithmic in nature, and gives identical statistical results for the analysis of variance.

Figure 2 shows the F0 pseudo-contours for the two-syllable utterances, and Figure 3, for the three-syllable ones. The Rise and Fall categories, which made up a large percentage of judgments for both languages, are clearly supported by the F0 contours. The only contour which is somewhat surprising is the two-syllable Fall pattern for the French-learning infants (Figure 2, top right). The second syllable starts high enough that, visually, we might be tempted to label this pattern a Rise/Fall. For the two-syllable complex patterns (middle panels of Figure 2), there does not appear to be enough change visible to base the judgment on. The three-syllable complex patterns, on the other hand, seem much easier to see. The flat patterns are truly flat (though there is only one per language in the three-syllable case). The Indeterminate patterns seem, for the most part, to have conflicting directions in the F0 changes, leading different listeners (indeed, the same listener on different occasions) to focus on one aspect or the other. In the two-syllable Indeterminate utterances, there is sizable variability, though not any larger than for some of the other categories. (The low variability in the three-syllable examples is due to low sample size.)

The F0 patterns support the listener judgments to a large degree, and most of the discrepancies can be attributed to differences in the analysis procedures. Our listeners were necessarily making prosody judgments, not F0 judgments, since the contribution of such factors as duration and amplitude cannot be avoided perceptually (Lehiste, 1970). Only three F0 measurements were taken per syllable, so some differences between syllables could have been lost. Some of those syllables might have been longer and therefore more important to the listener, but our analysis would be forced to ignore such a result. In the course of obtaining these medians, however, we did not feel that there were any instances of a contour within a syllable which was eliminated by being reduced to one number. Rather, the largest F0 excursions were reduced due to selection within a contour.

Another interesting perceptual effect that accompanied a few utterances was that of a large change in amplitude. While there were not enough such instances to study further, we discovered for them that a large decrement in amplitude was perceived as a drop in pitch, even when the acoustic analysis did not show a change in F0. Given that lower amplitude correlates well with decreased air flow, which in turn correlates with lower F0, this perception seems reasonable. Rossi (1979) has found just this correlation in the perception of steady-state vowels . It was not, however, taken into account in our acoustic analysis, and did not occur often enough to change the results.

Further analysis of the acoustic data was performed without regard to the perceptual results. We treated the F0 measurements within syllables and within utterances as repeated measures for the following analyses: comparisons of the first two syllables of all utterances; comparisons of the last two syllables of all utterances; and comparisons for the two- and three-syllable utterances separately of all the syllables in those utterances. Since it was not possible to combine the two- and three-syllable utterances directly, we compared syllables in the same relative position to the entire utterance in the first two subsets. These analyses showed language effects primarily in the interactions of Language with Syllable.


Rise

Fall

Rise/Fall

Fall/Rise

Flat

Indeterminate

Figure 2. **F0 patterns of the two-syllable utterances, categorized by the majority judgment made by the four listeners. The F0s are in averaged log Hz of the median F0 for each third of each syllable. The error bars represent one SD above and below that mean. The error bars for the French infants is always to the right of that for the English; if there are no error bars, the pattern represents only one utterance; there were no English-learning infant utterances which received a Fall/Rise majority. The log values for the tick marks from 2.3 to 2.8 correspond to 200, 251, 316, 398, 501, and 631 Hz.**

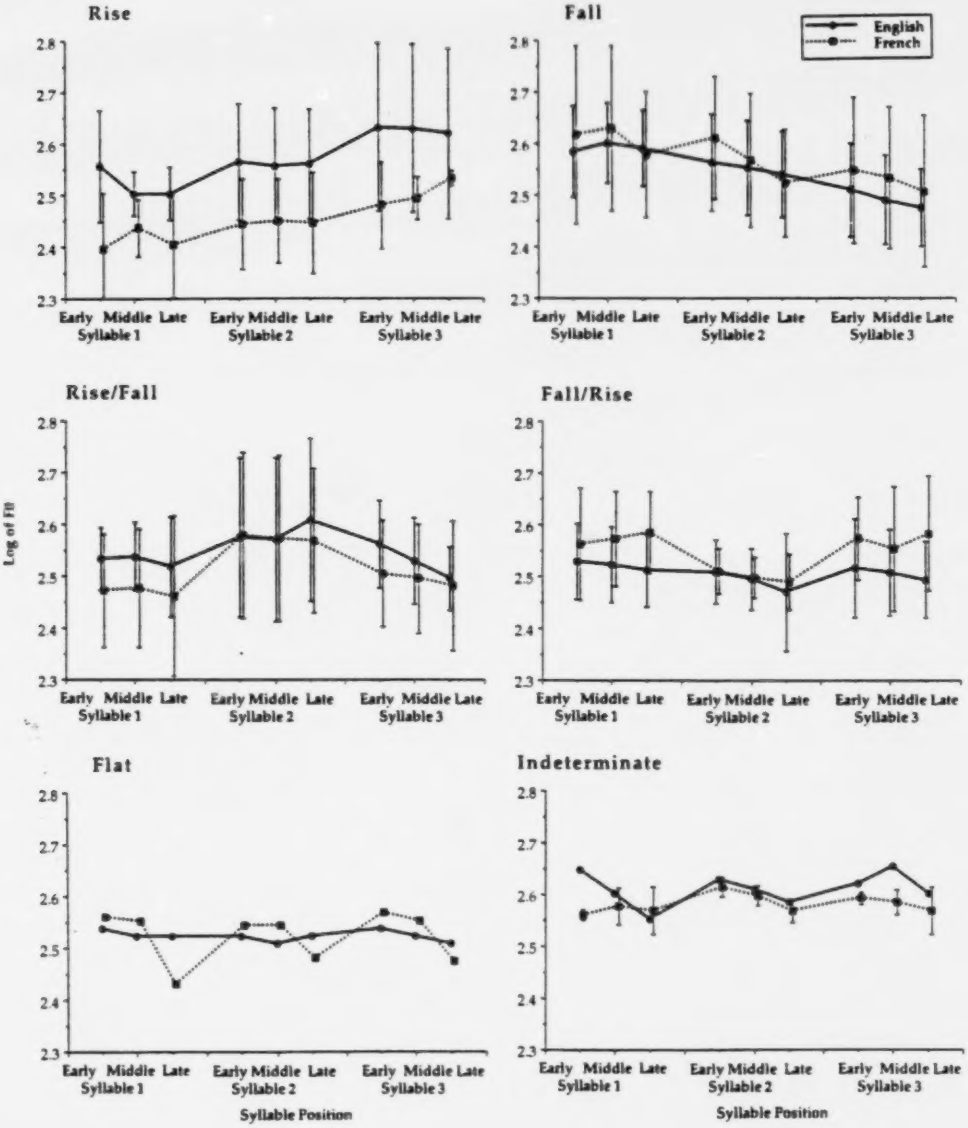


Figure 3. **F0 patterns of the three-syllable utterances, categorized by the majority judgment made by the four listeners. See Figure 2 for further details.**

In the analysis of the first two syllables of all utterances, the between factor of Language, and within factors of Syllable (first or second) and Location within syllable (early, middle late) were used. There was no main effect of Language or of Syllable (for both, F(1,154) < 1, n.s.), but a significant interaction between the two factors (*F*(1,154) = 6.91, *p* < .01). The English infants had higher F0s in the first syllable and lower in the second than the French, though the average of the two syllables was almost identical. Within syllables (Location), there was a consistent tendency for the F0 to drop (*F*(2,308) = 17.94, *p* < .001). The interactions of Location with the other two factors were not significant, showing that both syllables and both languages had the same tendency.

In the analysis of the last two syllables of all utterances, there was again no main effect of Language or Syllable ($F(1,154) < 1$, n.s.), but a significant interaction of Language with Syllable ($F(1,154) = 7.66$, $p < .01$). The English infants had higher F0s in the first syllable and lower in the second than the French, with the French two-syllable utterances with rising intonation contributed a great deal to that figure. The tendency for the F0 to drop within syllables was present as well in a significant Location effect ($F(2,308) = 26.03$, $p < .001$). The interactions of Location with the other two factors again were not significant.

The analysis of the two syllables of the two-syllable utterances revealed largely the same pattern. The Language and Syllable factors did not approach significance, but the Syllable by Language interaction was significant ($F(1,96) = 8.44$, $p < .01$). Location was also significant ($F(2,192) = 12.63$, $p < .001$). In this case, the Syllable by Location interaction was significant ($F(2,192) = 3.07$, $p < .05$), due to the second syllable's larger drop in value.

The analysis of the three syllables of the three-syllable utterances revealed a somewhat different pattern from that of the two syllable utterances. The Language factor was again not significant, though the Syllable factor did approach significance ($F(2,112) = 2.60$, $p < .10$). The Syllable by Language interaction was not significant for this analysis ($F(2,112) < 1$, n.s.). The Location changes were significant ($F(2,112) = 10.47$, $p < .001$). Not only was the Syllable by Location interaction significant ($F(4,224) = 2.46$, $p < .05$), but the three way interaction of Language, Syllable and Location was significant ($F(4,224) = 3.27$, $p < .05$). Each syllable had falling F0s, but the French utterances fell more steeply within each syllable, while the English utterances fell very little in the first syllable, but more steeply across the last two.

Discussion

Reduplicative babbles from five infants from a monolingual English environment and five from a monolingual French environment were analyzed for differences in intonation contour. The French children used significantly more rising intonation and less falling intonation than the English children, at least in the two-syllable utterances. This pattern reflects the relative distributions of F0 contours in the two adult languages (Delattre 1961). It also supports the findings of de Boysson-Bardies et al. (1984) that the language environment is discernable in the babbling of very young children.

The language-specific differences we have found could be confounded with the children's differential use of F0 for communication. D'Odorico (1984) found that even very young children (0;4-0;9) use different F0 contours in different proportions depending on their emotional state. Since we did not obtain detailed information on the context of each utterance, it was not possible to categorize the babbling on the basis of the child's presumed intent. Nonetheless, the recording situation excluded two of D'Odorico's three situations: Call and Discomfort were not relevant, since the parent was already present, and the children were recorded at the times they were least likely to be fussy. Even the number of requests (D'Odorico's third category) was likely to be small, since the parent was to interact with the child only when s/he stopped babbling. In another study of the communicative use of intonation contours, Marcos (1987) found that French children older than those studied here were likely to use rising contours for requests, and falling contours for labelling. However, the children in her study who were less than one and a half years old did not show this effect. Similar age ranges for the emergence of the communicative use of intonation has been found for English (Halliday, 1975; Menn, 1976). Thus it seems likely that the differences we found did not indicate that the French children were, by chance, making more requests. Rather, some aspect of the language environment seems to be responsible.

If the child learning a language is more influenced by the speech directed to her than by speech in the environment at large, then the differences in Infant Directed Talk (IDT) are of more direct concern than differences in the adult language. Despite the well-established increase in pitch changes in IDT relative to ADT (Adult Directed Talk) (e.g. Garnica, 1977), there is very little research examining pitch contours per se. For language directed at much younger infants, German appears to have a plurality of rising intonation (Fernald & Simon, 1984). For Mandarin, there is a plurality of falling contours (Grieser & Kuhl, 1988), though the infants in that study were much younger (0;2). The most extensive cross-linguistic study of IDT for children near the ages of the ones in this study does not report contours (Fernald et al., 1989). Fernald et al. do mention a common "tag vocalization" (1989:492) which has a steeply rising F0 for all the languages. At this point, we can only say that it is possible that the intonation that children between 0;5 and 1;1 hear is different in F0 pattern from the standard adult language, but a more detailed account must wait for further measurements. If the intonational differences are primarily present in the adult language rather than the more universal IDT, then the presence of intonational differences in babbling would signal an attunement to more than just IDT. Alternatively, such differences might ma ke us examine cross-linguistic differences in IDT more closely, in particular, the F0 contours. This is particularly important given the changes that occur in IDT as the child develops (Stern, Spieker, Barnett & MacKain, 1983; Bernstein Ratner 1984, 1986).

An apparently discrepant result in the literature is that of Allen (1983), who found that older French children (two-year-olds) had preponderantly falling intonation on their two- and three-syllable utterances. Allen's stimuli, however, were elicited by the presentation of a picture, which the child was supposed to name. Each utterance, then, was a complete statement of a sort, and as such would be expected, even in French, to have falling intonation. It is the rising intonation signalling continuation that is most different from English (Delattre, 1961), which tends rather to have a fall-rise.

Very little of the perceptually salient pitch change was present within single syllables. The two-syllable utterances received many fewer judgments of the complex patterns (Rise/Fall or Fall/Rise) than did the three-syllable ones. Other studies of children learning English have found a higher proportion of complex patterns (Kent & Murray, 1982; Robb, Saxman & Grant, 1989). In support of the present results, these earlier studies of English-learning children show a lack of rising contours similar to that found here. The difference in the percentage of complex patterns may be in part due to the way the contours were measured. The earlier studies used classifications of the F0 pattern based on visual estimation of the F0s from automatic pitch extraction. Since automatic pitch extraction algorithms make most of their mistakes at the boundaries between voiceless (or silent) and voiced segments, these systematic errors could introduce more apparent F0 movement than is actually present. Further, it is not obvious which patterns that make a visual difference will be perceived as distinct contours auditorily. The present methodology is more sensitive to the perceptually relevant changes and may therefore be the more relevant measure.

Our focus on reduplicative babbling, which was guided by methodological considerations, may have made it more difficult for F0 differences to emerge. Nonreduplicative babbling in the prelinguistic stage can contain long, continuous vocalizations with large changes in F0, but such utterances were deliberately excluded. The reduplicated babbling we examined is "poorer in prosodic cues" (de Boysson-Bardies et al., 1984:10) than babbling as a whole. That significant differences in F0 emerged nonetheless is even stronger evidence of the influence of the language environment.

One way in which languages differ is in whether F0 changes signal a change in words, as in a tone language such as Mandarin, or not, as in English and French. Since tones are characteristic of single words, it may be that the child learning a tone language will need to attend to F0 changes earlier than does the non-tone language learner. Also, the one-word stage may force the language learner to use an appropriately sized unit for a tone contour, while using a sentential intonation contour on a single word may be more difficult. This may explain the apparent earliness of tone learning relative to phoneme learning (Li & Thompson, 1976). One study of Mandarin IDT (Grieser & Kuhl, 1988) showed that mothers increase the global changes in F0 at the expense of tonal changes within the syllable, but that speech was directed at 2-month-old infants. Adults change their IDT as the infant changes (Bernstein Ratner, 1984, 1986), even in their range of F0 (Best & McRoberts, 1988). For a non-tone language, the meanings associated with these patterns are necessarily intonational. As such, they may be reflected in reduplicative babbling, which may be equivalent to a sentential construct and thus a site where native intonation could reasonably be attached. Further comparisons of the contours of late reduplicative babbling with the contours on early words would be of great interest. It would also be interesting to see the intersection of tone and intonation (cf. Lin, 1989) at the one word stage in a tone language.

The establishing of a difference in intonation between French and English sharpens the question of just what the language environment for the child is. It is clear that children prefer IDT to ADT (Fernald, 1985), but that does not indicate that IDT constitutes the major source of language exposure. If the contours used in IDT (or their relative frequency of occurrence) differ from those in the adult language, we would be able to determine somewhat better which style of language the child incorporates into her system. Similarly, it would be of interest to see if use of and sensitivity to F0 changes become language dependent before the well-established shifts in segmental perception

(Werker & Tees, 1984; Werker, 1989). Further work needs to examine the relative amount of time that infants are exposed to various language styles. From the present work, it appears that children are sensitive to F0 variation before they have a recognizable language system.

REFERENCES

Allen, G. D. (1983). Some suprasegmental contours in French two-year-old children's speech. *Phonetica, 40,* 269-292.

Bauer, H. (1988). Vowel intrinsic pitch in infants. *Folia Phoniatrica, 40,* 138-146.

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother-child speech. *Journal of Child Language, 11,* 557-578.

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics, 14,* 303-309.

Best, C. T., & McRoberts, G. W. (1988). Mutual adjustments of pitch by infant and caregivers during vocal interactions: A longitudinal case study. Paper presented at the International Conference on Infant Studies, Washington, DC.

Boysson-Bardies, B. de, Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language, 11,* 1-15.

Crompton, A. S. (1980). Intonation: tonetic stress marks versus levels versus configurations. *Word, 31,* 151-197.

Crystal, D. (1973). Non-segmental phonology in language acquisition: A review of the issues. *Lingua, 32,* 1-45.

Crystal, D. (1979). Prosodic development. In P. Fletcher & M. Garman (Eds.), *Language acquisition.* Cambridge: Cambridge University Press.

Cruttenden, A. (1986). *Intonation.* Cambridge: Cambridge University Press.

Delattre, P. C. (1961). La leçon d'intonation de Simone de Beauvoir, étude d'intonation déclarative comparée. *The French Review, 35,* 59-67.

D'Odorico, L. (1984). Nonsegmental features in prelinguistic communication: An analysis of some types of infant cry vocalizations. *Journal of Child Language, 11,* 17-27.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist, 66,* 103-114.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8,* 181-195.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10,* 279-293.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20,* 104-113.

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., Boysson-Bardies, B. de, & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16,* 477-501.

Garnica, O. K. (1977). Some prosodic and paralinguistic features of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition.* Cambridge: Cambridge University Press.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24,* 14-20.

Grover, C., Jamieson, D. G., & Dobrovolsky, M. B. (1987). Intonation in English, French and German: Perception and production. *Language and Speech, 30,* 277-294.

Kenning, M.-M. (1979). Intonation systems in French. *Journal of the International Phonetic Association, 9,* 15-30.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America, 72,* 419-425.

Leach, P. (1988). French intonation: Tone or tune? *Journal of the International Phonetic Association, 18,* 125-139.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

Lewis, M. M. (1936). *Infant speech: A study of the beginnings of language.* New York: Harcourt Brace.

Li, C. N., & Thompson, S. A. (1976). The acquisition of tone in Mandarin-speaking children. *UCLA Working Papers in Phonetics, 33,* 109-130.

Lin, H.-B. (1989). *The intersection of tone and intonation in Taiwanese.* Unpublished doctoral dissertation, University of Connecticut.

Locke, J. L. (1983). *Phonological acquisition and change.* New York: Academic Press.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7,* 41-75.

Marcos, H. (1987). Communicative functions of pitch range and pitch direction in infants. *Journal of Child Language, 14.* 255-268.

Menn, L. (1976). *Pattern, control and contrast in beginning speech: A case study in the acquisition of word form and functions.* Unpublished doctoral dissertation, University of Illinois.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation.* Unpublished doctoral dissertation, MIT.

Pike, K. L. (1945). *The intonation of American English.* Ann Arbor: University of Michigan Press.

Robb, M. P., Saxman, J. H., & Grant, A. A. (1989). Vocal fundamental frequency characteristics during the first two years of life. *Journal of the Acoustical Society of America, 85,* 1708-1717.

Rossi, M. (1979). Les configurations et l'interaction des pentes de F0 et de intonation. *Travaux de l'Institut de Phonetique d'Aix, 6,* 49-72.

Rossi, M. (1980). Prosodical aspects of speech productions. *Travaux de l'Institut de Phonetique d'Aix, 7,* 343-373.

Silverman, K. (1986). F0 segmental cues depend on intonation: the case of the rise after voiced stops. *Phonetica, 43,* 76-91.

Stern, D. N., Spieker, S., Barnett, R. K. & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language, 10,* 1-15.

Thevenin, D. M., Eilers, R. E., Oller, K. D., & Lavoie, L. (1985). Where's the drift in babbling drift? A cross-linguistic study. *Applied Psycholinguistics, 6,* 3-15.

Vihman, M. M., Ferguson, C. A., & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics, 7,* 3-40.

Vihman, M., Macken, M., Miller, R., Simons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language, 61,* 397-445.

Weir, R. H. (1966). Some questions on the child's learning of phonology. In F. Smith & G. A. Miller (Eds.), *The genesis of language* (pp. 153-168). Cambridge, MA: MIT.

Werker, J. F. (1989). Becoming a native listener. *American Scientist, 77,* 54-59.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Whalen, D. H., Wiley, E. R., Rubin, P. E., & Cooper, F. S. (1990). The Haskins Pulse Code Modulation (PCM) system. *Haskins Laboratories Status Report on Speech Research, SR-103/104,* 125-136.

FOOTNOTES

[*] *Journal of Child Language,* in press.

[†] Also Wellesley College.

[††] University of Connecticut, Storrs.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 41-62

From Babbling Towards the Sound Systems of English and French: A Longitudinal Two-Case Study*

Andrea G. Levitt† and Jennifer G. Aydelott Utman††

The utterances of one French and one American infant at 0;5, 0;8, 0;11, and 1;2 were transcribed and acoustically analyzed for syllable duration and vowel formant values. Both general and language-specific effects emerged in the longitudinal study. Initial similarities in the consonantal repertoires of both infants, increasing control in producing target F1 and F2 values, and developmental changes in babbling characteristics over time seem to reflect universal patterns. Yet the babbling of the infants differed in ways that appear to be due to differences in their language environments. Shifts in the infants' sound repertoires reflected phoneme frequencies in the adult languages. The English-learning infant produced more closed syllables, which is characteristic of English, than the French-learning infant. The French-learning infant tended to produce more regularly-timed nonfinal syllables and showed significantly more final-syllable lengthening, both characteristic of French, than the English-learning infant.

The study of very young children's acquisition of the sound patterns of their native language has focused, in particular, on whether or not there is continuity between babbling and early speech and on the roles of the linguistic environment and of general developmental processes in the child's early productions. Current research has produced much evidence in support of continuity between babbling and early speech and has suggested the importance of physiological maturation to the child's growing sound repertoire. However, empirical investigations of language-specific influences have had somewhat mixed outcomes, perhaps, in part, because of differences in the types of language-specific influences that were investigated (e.g., segmental versus prosodic) or in the techniques used (e.g., phonetic transcription versus acoustic analysis). Therefore, in our present investigation of continuity and language specificity, we both phonetically transcribed and acoustically analyzed a number of prosodic and segmental characteristics in the babbling development of an English-learning and a French-learning infant at four points between the ages of 0;5 and 1;2.

The best-known claim about the discontinuity between babbling and early speech comes from Jakobson (1968: 25). He suggested that 'the phonemic poverty of the first linguistic stages', as compared with 'the phonetic abundance of babbling', is indicative of a fundamental distinction between the two stages, and that the babbling period is therefore unrelated to the acquisition of language. He even proposed that infants pass through a silent period during the transition from babbling to speech, in which the child produces no sounds whatsoever.

This claim of discontinuity between babbling and early speech has not been supported by recent studies (Ferguson, 1978; Kent & Bauer, 1985; Menyuk, 1971; Menyuk & Menn, 1979; Oller, Wieman, Doyle, & Ross, 1976; Stark, 1980;

The research reported here was supported by grant DC00403 to Haskins Laboratories. Recording of the French infant was begun when the first author was a NATO postdoctoral fellow at the CNRS in Paris, France. Analyses of the data were conducted by the second author as part of a senior honors thesis at Wellesley College. We thank the families of our American and our French infants for their cooperation. We are grateful to the Sloan Foundation and to Wellesley College for additional support for this project. Portions of this paper were presented at the April 1990 conference of the International Conference on Infant Studies in Montreal, Canada. Catherine Best, Alice F. Healy, Michael Studdert-Kennedy, and Doug Whalen provided helpful comments on earlier drafts of the paper.

Vihman, Macken, Miller, Simmons, & Miller, 1985). For example, Vihman et al. (1985) examined the babbling and early word productions of nine children from 0;9 to 1;4 and found that babbling continued to occur after the onset of first words, without evidence of a silent period, and that the phonetic properties of babbled utterances both influenced and were influenced by word productions. Babbling increased in complexity following the acquisition of words, and some sounds appeared in babble only after words containing those sounds were learned. In addition, an infant's 'babble repertoire' was reflected in his or her later words.

The evidence that supports the role of physiological maturation in children's early vocal productions comes largely from studies that look at the developmental stages of babbling or at patterns in children's early segmental inventories. Oller (1980) and Stark (1980) defined several developmental stages of speechlike sound making within the babbling period that reveal babbling to be a logical and organized precursor to adult-like speech.

The first stage identified by both Oller and Stark begins at birth and lasts a month to six weeks. During this period, termed the 'phonation' or 'reflexive vocalization' stage, infants produce cry and fussing sounds as well as some partially resonant and a few fully resonant vocalizations. At the next stage, called the 'goo' or 'cooing' stage, roughly between 0;2 and 0;4, increasing vocal control gives rise to the infants' production of sounds that are typically heard as 'coos' by adults. At about 0;4 infants enter the 'expansion' or 'vocal play' stage, during which they produce a wide variety of vocalizations, including squeals, growls, and nasal murmurs. Towards the end of this period, they produce sounds that begin to approximate a number of adult consonants and vowels. Next, at about 0;6 or 0;7, infants begin a period of 'canonical' or 'reduplicated' babbling, in which they produce strings of identical consonant- and vowel-like elements, e.g., 'bababa.' The timing of these elements is much closer to that of adult speech than before. At the next stage, 'variegated' or 'nonreduplicative' babbling, which begins at about 0;10 or 0;11, infants are able to produce strings of syllables that contain a variety of consonantal and vocalic types, e.g., 'adega,' as well as a greater variety of syllable types. This last stage immediately precedes the infant's early word productions. Each stage generally reflects the onset and increasing incidence of a new type of vocalization, but the infant does not abandon earlier types of vocalizations. Thus, a child who is capable of variegated babbling may still produce utterances of the reduplicative type, just as a child who has begun to produce words also continues to produce babbled utterances.

Others have found additional evidence in support of the roles of continuity and maturational processes through the examination of infants' acquisition of phone types. Thus, Locke (1983) reviews the evidence for universal patterns in children's acquisition of speech sounds and argues in favor of anatomical and physiological maturation as the common source of these patterns.

Evidence for maturational processes has also emerged in studies of the acoustic properties of children's babbling and early speech. Buhr (1980) measured the vowel formant frequencies of the transcribed utterances of an American infant from 16 to 64 weeks of age and plotted the results on an F1/F2 plane. In describing the longitudinal development of the child's vowel space, Buhr attributes the early appearance of the sounds [i], [e], and [ɛ] to the early development of the control of jaw position. The sound [u], on the other hand, requires control of three elements: jaw position, which appears early on, longitudinal tongue movement, which arises early on in connection with swallowing, and lip rounding, which is also associated with sucking behavior. According to Buhr, the late emergence of [u] is due either to the difficulty of coordinating the three types of oral movement, or to the fact that voluntary control of lip rounding emerges considerably after its involuntary use in sucking. Finally, the development of neuromuscular ability in the lips, jaw, and front tongue is known to precede the physical descent of the larynx and rear tongue. Thus, Buhr attributes the emergence of the acute axis of the vowel triangle, which includes front vowels such as [i] , before the emergence of the grave axis, which includes the low back vowels such as [ɑ], to this difference in physiological development.

Other acoustic studies also chart basic developmental progress towards more adult-like productions. Bond, Petrosino and Dean (1982) measured the first and second formant frequencies of the speechlike productions of an infant from 1;5 to 2;2 and plotted the results on an F1/F2 plane. Their data suggest a refinement of the vowel space over time toward more adult-like formant frequency patterns, which presumably reflect the development of increased articulatory proficiency with age. Mack and Lieberman (1985) measured the babbling and words of an infant from 46-149 weeks for vowel length and consonant closure.

They found that specific segments of babble increased greatly in duration during the course of development, demonstrating the child's growing ability to approximate the longer adult-like norms.

But even though current research provides support for the ideas that there is continuity between babbling and speech and that maturation plays a role in the development of children's vocal productions, these factors do not preclude the possibility that the specific language environment may begin to influence vocalizations during infancy. The reports on language-specific influences have been mixed, however. A number of studies suggest an influence of the characteristics of the native language on the child's prelinguistic utterances, whereas others have failed to uncover language differences. One of the earliest suggestions for language-specific influences comes from Brown (1958: 199), who hypothesized that babbling 'drifts in the direction of the speech an infant hears', until the infant has mastered the adult phonology of the parents' language.

Some perceptual evidence appears to support Brown's notion of 'babbling drift.' Dinger and Blom (1973) asked 50 Dutch and 50 American women to listen to 90-second recordings of Dutch and American infants taken at 33, 43, 53, 62, and 70 weeks (approximately 0;8, 0;10, 1;0, 1;2, and 1;4), and to identify whether the babbling was Dutch or American. The women correctly identified the language background of the infants significantly often for ages 1;0, 1;2, and 1;4 months suggesting that after 1;0, language identification was possible. However, Dinger and Blom acknowledged that some words may have been included in the babbling samples, so that it is not clear on what basis the listeners made their judgments. Boysson-Bardies and her colleagues (Boysson-Bardies, Sagart, & Durand, 1984) also found that French adult listeners were able to distinguish at better than chance levels the babbling of French infants from the babblings of Chinese and Arab infants at 0;6 and 0;8, though they were no longer able to do so when listening to the babbling of ten-month-old infants (see also Weir, 1966). The stimuli consisted of 15 sec samples of babbling from a number of different children from each of the language backgrounds. Unlike the productions at 0;6 and 0;8, which had clearer melodic patterns, the stimuli from the productions of the 10-month olds were all highly articulated reduplicative babbles, and were thus poorer in intonation contour.

On the other hand, a number of other studies failed to find support for adults' ability to identify perceptually the linguistic background of the child from babbled productions (e.g., Atkinson, MacWhinney, & Stoel, 1970; Olney & Scholnick, 1976; Thevenin, Eilers, Oller, & Lavoie, 1985). Thus, Thevenin et al. (1985) examined the ability of adult monolingual speakers of English and bilingual speakers of English and Spanish to discriminate the babbling of English-learning from Spanish-learning infants at 0;7-0;10 and at 0;11-1;2. The stimuli consisted of 8 utterances each of 1- to 3-second stretches of canonical babbling from 14 infant subjects (7 English- and 7 Spanish-learning children). Adult listeners were not able to make correct language background identifications for the infants at either 0;7-0;10 or at 0;11-1;2, although they did show certain consistent patterns of response to particular infants and utterances.

One possible explanation for the discrepancy between the Thevenin et al. (1985) study and that of Boysson-Bardies et al. (1984) is that adults' ability to judge the linguistic background of a child may depend on the prosodic characteristics of the babbling, and thus may require long samples with pronounced melodic contour. It may also be that speech development, like the development of the child's phonological system, shows evidence of regression (Menn, 1971). Thus, children may in fact produce intonation contours that mirror those of the language of their environment at an early age and then apparently cease to produce such patterns consistently as they turn their attention to the production of more segmentally complex utterances. It is then perhaps not surprising that the French judges in the Boysson-Bardies et al. (1984) experiment were unable to identify the language background of the 10-month-old children, although they had done so for the younger children, because the prosodic cues were apparently poorer in the productions of the 10-month olds.

More recently, however, studies of the acoustic and phonetic properties of infants' babbling from specific language backgrounds have provided evidence in support of the babbling drift hypothesis. Boysson-Bardies, Halle, Sagart, and Durand (1989) found that the distributions of vowel formant frequencies of five 10-month-old infants each from French, English, Cantonese, or Arabic backgrounds reflected the different formant-frequency pattern distributions of adult vowels in the target languages, a result that demonstrates the early influence of the language environment on vowel production in infant babbling.

Similarly, Boysson-Bardies (1989) reports evidence of babbling drift in a cross-linguistic study involving infants from French, American, Japanese, and Swedish backgrounds. Five children from each linguistic community were recorded in six sessions, starting at 0;10, when they could only babble, and continuing until the 25-word stage. In this study, transcriptions of the consonantal repertoires of the infants showed evidence both of universal developmental processes and of language-specific influence from about age 0;10 on. Thus, stops represent, by and large, the largest proportion of consonants for infants from all language backgrounds, presumably because they are easier to produce than other consonants. However, from the session at 0;10 onwards there was a clear stop frequency difference between the French infants, on the one hand, and the American and Swedish infants, on the other. The latter groups regularly produced a higher mean percentage of stops than did the French. This difference is consistent with the higher frequency of stop consonants in the target-words attempted by the American infants and those attempted by the Swedish infants when compared to the frequency of stop consonants in the target words attempted by the French infants.

Most studies examining the effect of the linguistic environment have generally taken a cross-sectional rather than longitudinal approach. On the other hand, most of the previous research investigating the role of maturational factors in speech development has focused on a single language longitudinally. Even in the rare cases where a study of the effect of the linguistic environment was longitudinal, as in the case of Boysson-Bardies (1989), the data were usually only phonetically transcribed and not acoustically analyzed. If we are to determine the time course of maturational and language-specific factors on speech development, we need to make longitudinal cross-linguistic comparisons. Furthermore, acoustic analyses may reveal effects that are not apparent in perceptual or transcription studies. In an acoustic analysis of voice onset time (VOT), Macken and Barton (1980) found that some children, who were heard by adults to be using initial voiced stops for voiceless targets, were actually producing longer VOTs for their voiceless targets than for their voiced ones. The difference between the two types of stops was, however, inaudible to the unaided ear. Therefore, the present study examined the role of maturational and language-specific effects longitudinally by comparing both the transcriptions and the acoustic analyses of the vocal productions of a French-learning and an American English-learning infant at 0;5, 0;8, 0;11, and 1;2. A case study approach to this problem is appropriate as a means of generating specific hypotheses about the speech development of French- and English-learning infants that can later be tested on a larger sample of subjects from each language background.

French and American English infants were chosen for comparison because of specific known segmental, syllabic, and prosodic differences between French and English. A comparison of the sound inventories of the two languages reveals that English has diphthongs, which are not found in French, more fricatives and affricates than French, and one more nasal consonant than French (/ŋ/), while French, on the other hand, has nasal vowel phonemes, which are not found in English (Delattre, 1965). Syllabically, the two languages also differ in the most common syllable structures and the typical word lengths (in numbers of syllables). English contains more closed syllables and more monosyllabic words than French (Delattre, 1965). Finally, in terms of prosody, French and English differ in fundamental frequency (F_0) contours and durational patterns. For the present study, we focused on acoustic measures of durational pattern differences between the adult languages. French and English are classified respectively as syllable-timed and stress-timed languages (e.g., Pike, 1945, but cf. Wenk & Wioland, 1982). French, the syllable-timed language, has a rhythmic structure characterized by nonfinal syllables generally equal in length, or isosyllabicity, and by regular breath-group-final lengthening. English, the stress-timed language, has phrase-final lengthening, but this interacts with lengthening due to lexical stress, and therefore does not show the relative nonfinal isosyllabicity of French.

Therefore, if the language environment influences infant babbling, we should expect to see greater nonfinal isosyllabicity, more phrase-final lengthening, and perhaps more polysyllabic utterances in the babbling of the French infant, whereas we should see more monosyllabic productions and more utterances with closed syllables in the babbling of the American infant. Furthermore, if the infants' productions are affected by the segmental inventories of their linguistic environments, their productions should, over the course of time, reflect that influence in the proportions of specific vowel and consonant types. Thus, language-specific effects were assessed through the examination of segmental features, syllable type, and timing, both in terms of the transcriptions of the utterances (consonant and vowel inventories) and in terms of their acoustic properties as well (the F1 and F2 formant frequency values of the vowels and durational measures of the syllables).

General effects, or language-independent similarities in the developmental patterns of the two infants, should be revealed through the study of syllable types and segmental features. The babbling sequences produced by both infants over time should evolve in keeping with the stages described by Stark (1980) and Oller (1980). If physiological constraints are the main determinants of the types of sounds that very young infants produce, the vocal repertoires of the two infants should be more similar to one another at earlier stages rather than at later stages, when other determinants (from the surrounding language) have also come into play. Finally, over time, both infants should improve in their ability to produce the target formant frequency values of their vowels and thus show less variability in those measurements.

Method

Subjects. The babbling of a male American English-learning infant (E-CR) and that of a female French infant (F-YC) were recorded in their homes by their parents on a weekly basis from 0;5 to 1;5. Both infants were second children, each with a single sibling of the same sex at least four years older. The American infant, recorded in a household in Providence, RI, heard regionally typical American English and the French infant, recorded in Paris, France, heard Parisian French.

Materials. E-CR was recorded using a Panasonic cassette tape recorder with a Realistic supercardioid 33992A microphone. YC's utterances were recorded using a Panasonic RQ 3145 cassette tape recorder with a high-quality cardioid microphone.

Procedure

Data collection. Individual recording sessions lasted approximately 10-20 minutes. The parents were instructed to choose a time for recording when the infant was likely to be alert and unlikely to cry (e.g., in the morning after eating). The parents were to hold the microphone at about 20 cm from the baby, avoiding external noise, and to say the date at the beginning of each recording to identify individual sessions. The parents could either elicit babbling by talking and gesturing (making sure to stop talking while the baby vocalized) or record spontaneous babbling as it occurred. A comment sheet was filled out for each tape, including date, time, and situation (e.g., 'in bath' or 'watching sister'), and the parents were asked to indicate whether babbling was spontaneous or elicited by a specific person.

Utterance sampling. The utterances used in this analysis were taken from four separate months of recordings beginning when the children turned 0;5, 0;8, 0;11 and 1;2. These ages were chosen to capture four of the stages of infant vocalization that were described earlier: vocal play, reduplicative babbling, variegated babbling, and the period of the onset of the child's first words (Oller, 1980; Stark, 1980). Only utterances containing V, CV, VC and CVC syllables with vocalic durations greater than 100 ms were selected for analysis. No other productions typical of the vocal play period (i.e., squeals, growls, etc.), emotive sounds (i.e., crying, laughing, etc.) or vegetative noises were used.

Phonetic transcription. Utterances were divided into separate strings or breath groups, which were defined as an individual syllable or sequence of syllables separated from other utterances by a silent interval of 400 ms or more. Each utterance was transcribed by two phonetically trained listeners, both of whom are fluent in English and French. Utterances for which the transcribers disagreed were retranscribed until agreement was reached.[1]

Acoustic analysis. The utterances were input at half speed and digitized with a Macintosh Plus computer at a sample rate of 10 KHz, using a MacADIOS digitizer and a TTE 411AFS amplifier. The utterances were measured for vocalic F1 and F2 frequencies (Hz) at the peak amplitude portion of each vowel. Vowels for which either F1 or F2 measurements were indeterminate were dropped from the formant frequency analysis. For F-YC, 15.9% and for E-CR, 15.8% of the vowels overall gave indeterminate measures. Eleven utterances were randomly selected from each infant's productions for remeasurement. For E-CR these utterances contained a total of 41 syllables, and the overall mean discrepancy for F1 and F2 measurements was under 15 Hz. For F-YC these utterances contained a total of 69 syllables, and the overall mean discrepancy for F1 and F2 measurements was under 20 Hz.

Syllabic durations (ms) were also measured. In multisyllabic utterances, measurements were made by visually inspecting the waveform and by listening to each of the measured portions. A conservative criterion for measurement of vowel-final syllables was adopted: only visibly voiced portions of the final syllable were included. Remeasurement of the durations of the same random sample of E-CR's utterances was within 23 ms on average, while remeasurement of the durations of F-YC's eleven utterances was within 13 ms on average. A total of 881 syllables for the French infant (114, 317, 201, and 249 syllables at 0;5, 0;8, 0;11, and 1;2, respectively) and 701 syllables for the American infant (178, 218, 218, and 87 syllables at 0;5, 0;8, 0;11, and 1;2, respectively) were measured.

Results

The results will be described in terms of segments, syllable structure, and timing, in that order. The transcriptions provide the data for an analysis of consonant and vowel frequencies of occurrence and of the syllable structure. The acoustic analyses provide the data for formant analyses of the vowels and for the timing characteristics of the children's utterances.

Segmental Features

Transcription analysis. The sound systems of English and French will be discussed first in terms of the consonant and vowel frequencies and then in terms of the distribution of consonants according to manner and place of articulation. In order to examine longitudinal patterns of change in segment frequencies, charts of percent distribution of vowels and consonants were created for each infant at each age and then listed in order of percentage of occurrence (see Tables 1-4). The charts were compared to lists of the sounds of French and English in order of percentage of occurrence (see Table 5; Delattre, 1965).

As can be seen from Table 5, English and French share 18 consonants of the 25 listed. Six consonants appear in English but not in French (generally fricatives and affricates and the nasal /ŋ/), and one appears in French but not in English (/ɥ/, a labial-palatal approximant). Of the 21 vowels, 8 are shared by the two languages, 5 occur in English but not French (mostly diphthongs) and 8 occur in French but not English (front rounded and nasal vowels).

Across all ages observed, the infants had 27 segments in common in their productions, and each had an additional 8 sounds that the other did not use. Sixteen of the common segments were consonants, thirteen of which appear in both French and English. The English-learning infant produced an additional 7 consonants, mostly fricatives and affricates, two of which are only found in English ([ʤ] and [ʧ]), and one ([x]) that is found in neither English nor French. The French infant produced only the shared 16 consonants, three of which occur only in English and not in French. As for the vowels, both infants produced the same 11 vowels, though with different frequencies of occurrence. The eleven vowels included two non-French vowels ([æ] and [ɪ]) and one non-English vowel ([ø]).

Table 1. *List of consonants in order of percentage of occurrence at 0:5, 0:8 , 0:11, and 1;2 for E-CR (Percentage base: total number of consonants in syllables analyzed at each age: 178, 218, 218, and 87 at 0;5, 0;8 , 0;11, and 1;2, respectively).*

0;5		0;8		0;11		1;2	
x	24.4	b	30.4	d	17.0	d	27.4
ɡ	15.9	w	15.9	ʔ	14.8	ʃ	25.8
h	14.8	h	10.9	w	8.2	ʧ	11.3
w	12.7	j	8.9	ʤ	7.7	b	8.1
ʔ	11.7	ɡ	6.5	j	7.7	ʤ	8.1
j	9.5	v	6.5	h	7.1	j	6.5
m	3.2	x	6.5	ʃ	6.6	ð	3.2
l	3.2	l	5.4	b	6.0	w	3.2
f	2.1	ʔ	3.3	l	4.4	f	1.6
p	1.0	d	2.2	t	3.3	h	1.6
v	1.0	ʤ	1.1	ɡ	3.2	s	1.6
		m	1.1	n	2.7	t	1.6
		ð	1.1	ʧ	2.2		
		z	1.1	s	1.6		
				ʒ	1.6		
				ð	1.1		
				v	1.1		
				x	1.1		
				f	0.5		
				m	0.5		
				r	0.5		
				z	0.5		

Because French and English share many of the same segments, it is not clear whether the infants' productions are similar because of the similar sounds to which they were exposed, or because, as Locke (1983) would argue, certain sound patterns are particularly common in the languages of the world because they reflect both the child's and the adult's natural phonetic proclivities. However, an examination of the longitudinal changes in each infant's productions suggests target-language influence.

Most of the consonants in the American infant's babble repertoire (see Table 1) which showed an increase in relative frequency over time ([d, ð]) or which began to appear at 0;11 ([t, n, r, s, z]) are more frequent in the English language than those which diminished in relative frequency ([b, h, w, j]) or disappeared altogether from the inventory ([ɡ, p, v]). The American infant, CR, also introduced or increased the relative frequency of voiced and unvoiced alveolar and palato-alveolar fricatives and affricates ([ʧ, ʤ, ʃ, s, ʒ]) at 0;11. It is possible that CR developed the neuromuscular ability to produce these sounds at about this time, resulting in an increase in their production. The only non-English sound in CR's inventory is [x], which showed a gradual decrease over time and disappeared entirely by 1;2. There are only two consonants in English (/ŋ/ and /θ/) that CR did not produce.

CR's vowel productions (see Table 2) show a similar effect. CR displayed a preference for [ə], [æ], and [a] at all months, and percentages for each remained relatively stable over time, consistent with their high frequency of occurrence in English. The sound [e] was relatively infrequent in CR's babble repertoire longitudinally, and is also infrequent in English. Dramatic increases in the relative frequency of [ɪ] and [ɛ] at 0;11 and, at 1;2, increases in the frequency of [i] and decreases in [ɔ] and [u] are also consistent with the relative frequency of occurrence of these sounds in English.

Table 2. *List of vowels in order of percentage of occurrence at 0;5, 0;8 , 0;11, and 1;2 for E-CR (Percentage base: total number of vowels in syllables analyzed at each age: 178, 218, 218, and 87 at 0;5, 0;8, 0;11, and 1;2, respectively).*

0;5		0;8		0;11		1;2	
ə	38.1	ə	27.9	ə	31.1	ə	36.4
æ	19.0	a	21.4	a	27.4	æ	17.0
u	13.5	u	20.0	ɛ	9.0	a	12.5
ɔ	9.2	æ	11.6	ɪ	8.0	ɪ	9.1
a	8.0	ɛ	7.4	æ	5.7	ɛ	9.1
ɛ	4.3	i	3.3	ɔ	5.2	i	5.7
i	1.8	e	3.3	i	3.8	e	3.4
ɪ	1.8	ø	2.8	u	3.8	ɔ	2.3
o	1.8	ɪ	1.9	e	1.9	o	2.3
e	1.2	ɔ	0.5	o	1.9	u	2.3
ø	1.2			ø	1.9		

Similar language-specific patterns are also apparent in the repertoire of the French infant, YC. YC showed a definite preference after 0;8 for the sound [d], which has a relatively high frequency of occurrence in French (see Table 3), but this sound is also frequent in English and occurred often in CR's productions as well. Nonetheless, in general, most consonants which showed increased frequency of occurrence longitudinally ([b, m, n, j]) or appeared in the inventory at 1;2 ([p]) occur more frequently in French than the consonant which decreased in frequency ([ɡ]). In addition, YC initially produced a number of non-French sounds in her inventory, but these sounds diminished greatly in frequency by 1;2 ([h]), disappeared gradually over time ([ð]), or appear only at 0;5 or 0;8 without recurrence ([ŋ, ʔ, r, x]). YC did not display the increased production of alveolar fricatives and palato-alveolar fricatives and affricates shown by CR at 0;11, although [ʤ] and the palatal glide [j] did appear in her inventory at about this time. However, affricates do not exist in French, and [ʤ] did not reappear in YC's speech sample at 1;2.

Table 3. *List of consonants in order of percentage of occurrence at 0;5, 0;8 , 0;11, and 1;2 for F-YC (Percentage base: total number of consonants in syllables analyzed: 114, 317, 201, and 249 at 0;5, 0;8 , 0;11, and 1;2, respectively).*

0;5		0;8		0;11		1;2	
l	36.6	d	66.4	d	58.4	d	30.8
d	29.3	l	12.6	j	10.3	b	16.8
h	18.3	m	5.3	w	9.9	j	14.4
ɡ	3.7	h	3.2	b	8.1	m	9.6
ŋ	3.7	j	3.2	l	5.4	n	8.2
m	2.4	n	2.4	ˈj	3.2	p	7.7
b	1.2	b	1.6	ð	2.2	l	5.8
r	1.2	ɡ	1.6	h	1.1	w	3.4
t	1.2	t	1.2	n	1.1	t	1.4
ð	1.2	ð	1.2	m	0.5	ɡ	1.0
v	1.2	w	0.4	t	0.5	h	1.0
		x	0.4				
		ʔ	0.4				

Table 4. *List of vowels in order of percentage of occurrence at 0;5, 0;8 , 0;11, and 1;2 for F-YC (Percentage base: total number of vowels in syllables analyzed: 114, 317, 201, and 249 at 0;5, 0;8 , 0;11, and 1;2, respectively).*

0;5		0;8		0;11		1;2	
æ	44.2	a	39.6	æ	40.5	a	33.5
a	26.5	æ	27.7	i	14.1	æ	24.6
ə	19.5	ə	17.3	ə	11.2	ə	19.8
ɛ	2.7	ɛ	4.5	a	10.7	u	8.9
e	1.8	ø	3.0	ɛ	7.8	ø	5.6
u	1.8	u	2.4	e	4.4	ɪ	2.0
ø	1.8	ɪ	2.1	u	3.4	i	1.6
i	0.9	o	2.1	ɪ	3.0	ɛ	1.6
ɪ	0.9	i	1.2	ø	3.0	e	0.8
		ɔ	0.3	o	2.0	ɔ	0.8
						o	0.8

In general, YC produced the same vowels as CR, in terms of the transcriptions assigned them. Her preference for low front vowels ([a] and [æ]) was stable at each of the four months sampled. However, the mid central vowel consistently favored by CR ([ə]), which is the most frequent vowel in English (see Table 5), remained in third place for YC, and is the fourth most frequent vowel in French. The French sounds /y/ and /ɑ/ and the French nasal vowels did not occur in YC's productions. It is possible that these sounds failed to emerge, and that the children's vowel inventories were similar, because of physiological constraints on vowel production, such as inability to produce the necessary voluntary lip-rounding or the appropriate lowering of the velum.

Table 5. *Lists of consonants and vowels in order of percentage of occurrence in English and French (Delattre, 1965 (Percentage base for vowels: total number of vowels in the sample for the relevant language. Percentage base for consonants: total number of consonants in the sample for the relevant language). Sounds not shared by the two languages are indicated by underlined type.*

English				French			
consonants		vowels		consonants		vowels	
t	12.7	ə	24.1	r	15.0	e	19.3
n	11.4	ɪ	15.1	l	10.6	a	16.7
r	8.3	æ	9.9	t	9.7	i	12.4
l	7.6	i	8.9	s	8.8	ə	7.6
s	7.4	a	7.3	p	8.0	ɑ̃	7.6
d	6.1	ɛ	7.2	d	7.2	ɛ	6.7
z	4.9	u	5.9	k	6.4	u	6.4
m	4.7	ai	5.7	m	6.0	ɔ	5.1
ð	4.6	o	5.2	n	5.2	y	4.7
k	4.3	e	4.1	v	4.5	ɔ̃	3.8
w	3.7	au	2.3	j	3.2	o	2.6
b	3.4	ɔ	2.1	ʒ	3.0	ɛ̃	2.4
h	3.3	ʊ	2.1	f	2.6	œ	1.8
v	3.2	oi	.2	z	2.3	ø	1.7
f	2.8			w	2.3	œ̃	1.0
p	2.3			b	2.3	ɑ	.1
ŋ	2.2			g	1.1		
j	2.0			ʃ	.9		
g	1.6			ɥ	.8		
θ	.9						
ʃ	.8						
dʒ	.8						
tʃ	.6						
ʒ	.2						

Patterns of change in relative frequency of occurrence for vowel sounds are not as clear for YC as they are for CR, with the exception of an increase in the production of [u] over time, a decrease in [o], and the stable high frequency of [a] and [ə], all of which correspond with the relative distributional frequency of each sound in the Delattre data.

Overall, these longitudinal patterns of production, for both infants, particularly in the case of consonants, seem to suggest the possibility of transition into the adult language. Indeed, when we calculate the percentage of the total consonantal inventory at 0;5, 0;8, 0;11, and 1;2 that the children share, we see that it starts at 35% at 0;5, increases to 69% at 0;8 and then diminishes to 50% at 0;11 and 35% again at 1;2. This gradual decrease from a high of 69% overlap at 0;8 and each child's increased production of the consonant sounds in the ambient language suggest language-specific developmental effects.

Some further support for language-specific patterns emerges if we consider the children's consonant productions at each age in terms of manner (Figure 1) and and their vowel productions in terms of vowel height and frontness (Figure 2). Each figure also contains the data for the appropriate adult language from Table 5 collapsed into the same categories.

As can be seen from Figure 1, English uses fricatives, affricates and nasals more than French, whereas approximants are more frequent in French. Surprisingly, both infants showed perhaps their closest match to the frequency distributions of the manner categories of their native language in their vocal productions at 0;5 (Figure 1). Both infants also showed a strong preference for producing stops at 0;8 and 0;11, though more so in the case of the French infant YC. At 1;2, the infants appeared to be heading back towards a pattern that closely matches that of their native language, except that the French infant, YC, produced very few fricatives, and the American infant, CR, produced somewhat more fricatives and affricates than stops (as he did at 0;5) and very few nasals. The results for CR at 1;2 months are based on the smallest number of syllables (87), and should perhaps be regarded with caution.

The early preference for stops on the part of both infants is consistent with the universal tendency for early consonant productions to be stops (Locke, 1983). Although our French infant produced more stops, consistent with their distribution in adult French, than our American infant, Boysson-Bardies (1989), on the other hand, found that French infants produced fewer stops than American infants. However, her study began with older infants (0;10), followed them through the 25-word stage, and compared the children's distribution of consonants with the consonants in the adult models of the children's target-words rather than the language as a whole. The discrepancy may thus be due to the fact that our two infants were largely prelinguistic,[2] whereas those of Boysson-Bardies had made entry into the linguistic systems of their environments.

In the case of the vowel categories, the infants' productions were divided into seven categories, front high (fh), front mid (fm), front low (fl), mid central (mc), back high (bh), back mid (bm) and back low (bl). As can be seen from Figure 2, F-YC favored front vowels, in particular, front low, which is consistent with the higher frequency of front as compared to central and back vowels in French. On the other hand, E-CR produced considerably more mid central vowels, which is consistent with their higher frequency in English. Furthermore, over the period of investigation, E-CR showed an increase in front high vowels, which is also consistent with their relatively high frequency in adult English.

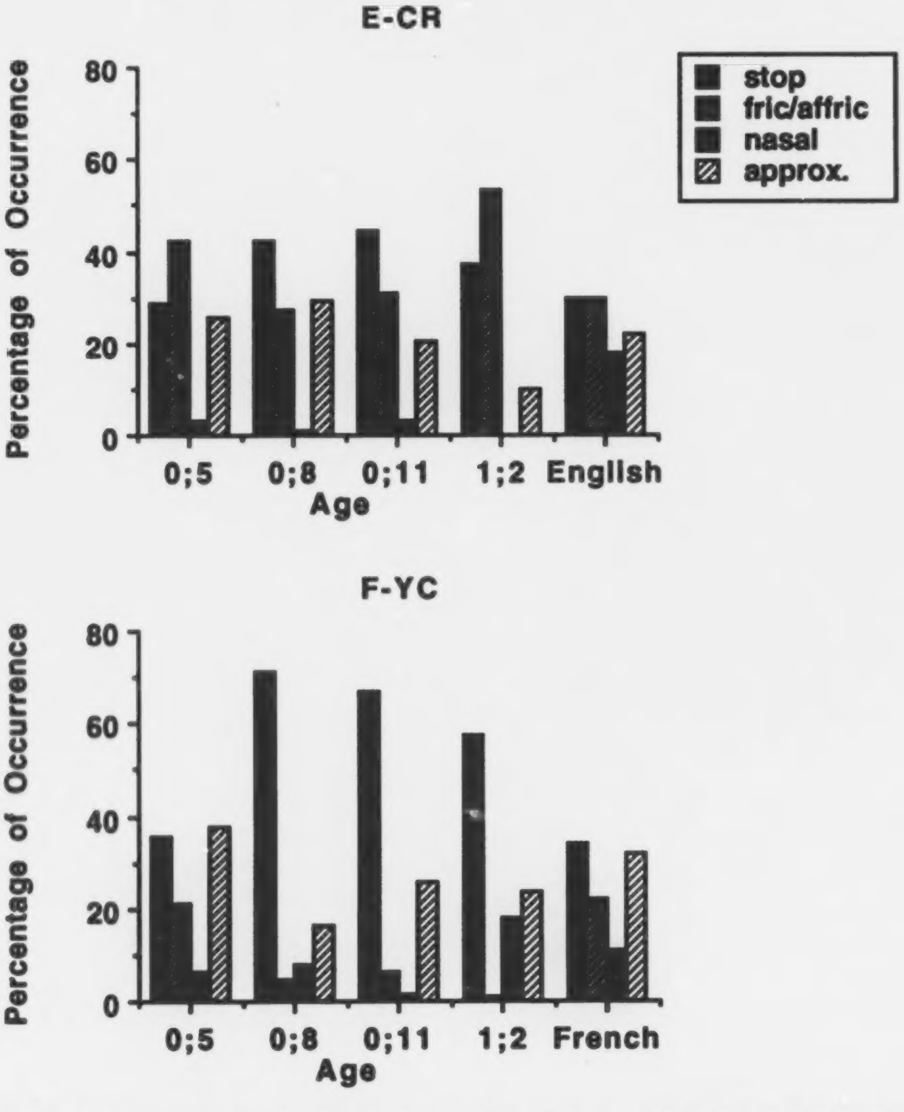


Figure 1. Distribution of the consonant productions of the American (E-CR) and the French (F-YC) infant as a function of manner class at 0;5, 0;8, 0;11, and 1;2. [ʔ] and [h] were not included in the graph. The bars at the right of each graph indicate the manner class distribution of consonants for each of the adult languages.

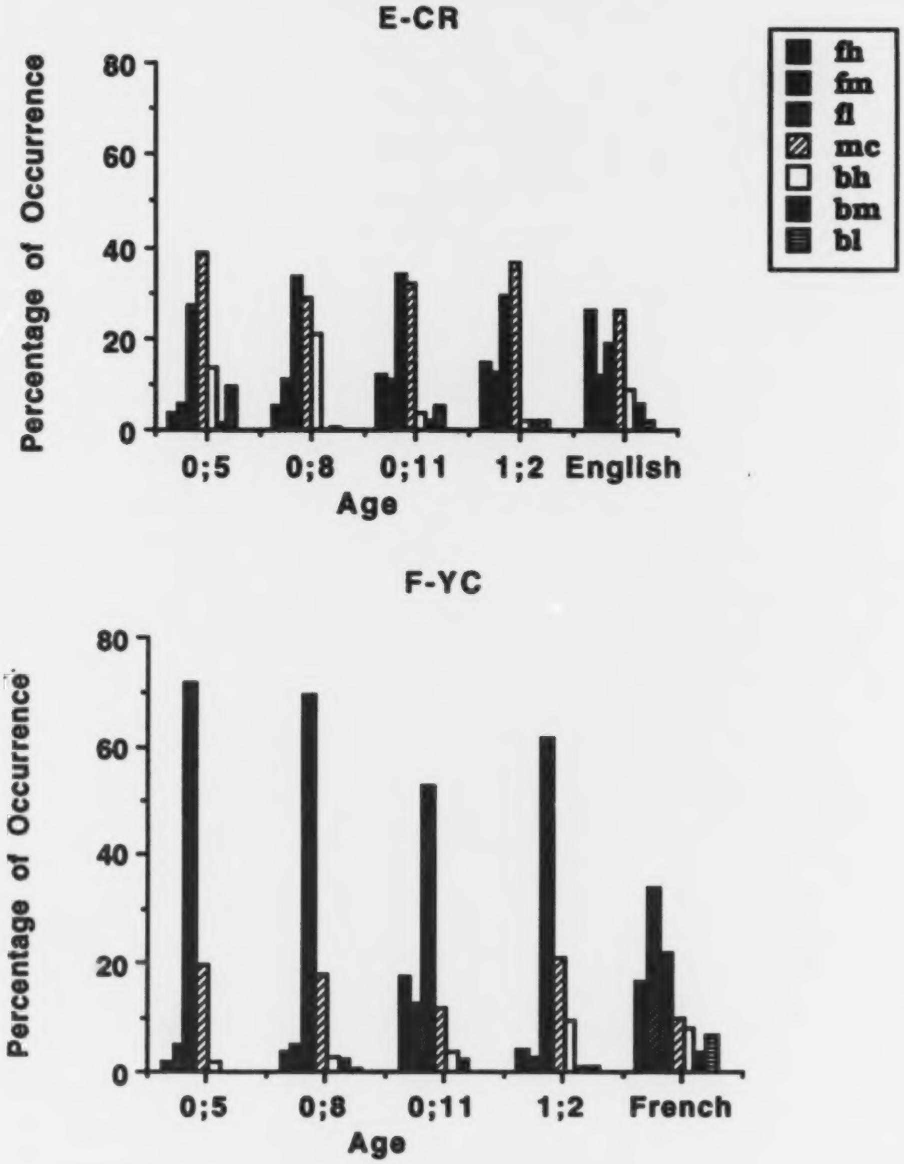


Figure 2. **Distribution of the vowel productions of the American (E-CR) and the French (F-YC) infant as a function of seven categories, front high (fh), front mid (fm), front low (fl), mid central (mc), back high (bh), back mid (bm) and back low (bl) at 0;5, 0;8, 0;11, and 1;2. The bars at the right of each graph indicate the vowel category distribution for each of the adult languages.**

The percentage of each infant's consonant and vowel productions in terms of the categories established in Figures 1 and 2 was correlated with the percentages of those categories in the adult languages. Table 6 shows the results of those correlations. Most of the correlations themselves are not significant (except one for F-YC at 0;5 and two for E-CR at 1;2), as would be expected given the fact that the infants' consonantal and vowel repertoires are more limited than the adults' because of developmental constraints. However, it is striking that in 14 out of 16 comparisons ($p <$ 0.002, by a binomial distribution test), the French infant's productions correlate more strongly than the American infant's productions with the French consonant and vowel categories, whereas the

American infant's productions correlate more strongly than the French infant's productions with the English consonant and vowel categories.

Table 6. *Correlations between the infant consonant productions and those of English and French in terms of manner categories and between infant vowel productions and those of English and French in terms of vowel height and frontness.*

	French Consonants		English Consonants	
	F-YC	E-CR	E-CR	F-YC
Age				
0;5	0.99*	0.56	0.88	0.47
0;8	0.67	0.94	0.81	0.49
0;11	0.81	0.78	0.94	0.52
1;2	0.61	0.34	0.96*	0.23
	French Vowels		English Vowels	
	F-YC	E-CR	E-CR	F-YC
Age				
0;5	0.32	-0.03	0.51	0.37
0;8	0.31	0.26	0.53	0.36
0;11	0.50	0.32	0.72	0.50
1;2	0.25	0.31	0.82*	0.39

*$p < 0.05$

Formant analyses of the vowels. Mean F1 and F2 frequencies and standard deviations for all vowels at each age were plotted for both infants (see Figure 3) and compared with the findings of Boysson-Bardies, Halle, Sagart, and Durand (1989) for English and French infants. Of the 178, 218, 218, and 87 syllables analyzed at 0;5, 0;8, 0;11, and 1;2, respectively, for E-CR, 26, 38, 33, and 14 syllables gave indeterminate F1 or F2 measurements and were excluded from the analyses. Similarly, of the 114, 317, 201, and 249 syllables analyzed at 0;5, 0;8, 0;11, and 1;2, respectively, for F-YC, 9, 31, 35 and 65 syllables at each of the four ages gave indeterminate F1 or F2 measurements and were excluded from the analyses. Mean F2 for the English-learning child, E-CR, remained relatively stable until 0;11 when it began to increase. Mean F2 for the French-learning child, F-YC, remained relatively stable until 1;2 when it began to decrease. At 1;2 the mean F2 for E-CR was higher than that of F-YC. This pattern is consistent with the higher mean F2 found by Boysson-Bardies et al. (1989) in the productions of English-learning infants when compared to those of French-learning infants, although this difference appeared in their data at 0;10.

Mean F1 remained fairly stable for both children, with the mean F1 value of the American child, E-CR, increasing slightly over the four months sampled, and the mean F1 value of the French child, F-YC, decreasing slightly over the same period. Overall, the mean F1 was lower for E-CR than for F-YC, but the average values appear to be converging, while in the case of F2, they seem to begin to diverge. Boysson-Bardies et al. found essentially identical measurements for mean F1 for their sample of English and French infants, so this convergence is perhaps to be expected. Although language-specific differences occurred later in this analysis than in that of Boysson-Bardies et al., they nonetheless appear to suggest the influence of the language environment on the acoustic properties of vowels for both infants.[3]

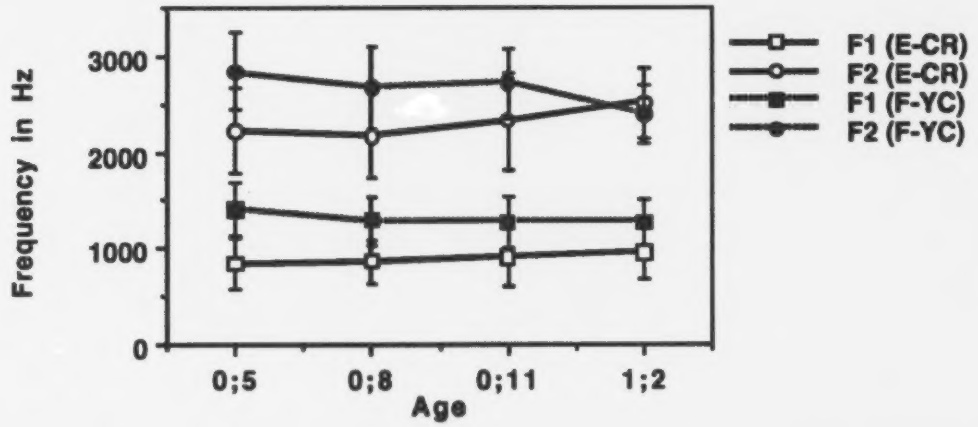


Figure 3. Mean F1 and F2 and standard deviations for the American (E-CR) and the French (F-YC) infant at 0;5, 0;8, 0;11, and 1;2.

Boysson-Bardies et al. (1989) have argued that from about 0;10 of age, infants have sufficient articulatory control to produce vocal tract positions that result in sound productions that reflect the ambient language. We explored the possibility of such an influence by examining the formant frequency patterns for ten vowels produced by the two infants. Mean and standard deviation of F1 and F2 for the ten vowel sounds were plotted for each infant at each age (see Figures 4 & 5). In the adult languages, F1 generally varies inversely with vowel height, whereas F2 is generally higher for front vowels and lower for back vowels. F1 can thus often be correlated with correct jaw control (or tongue height), whereas F2 usually corresponds to tongue fronting (but cf. Ladefoged (1975) for some qualifications on inferring tongue position from F2 for back vowels).

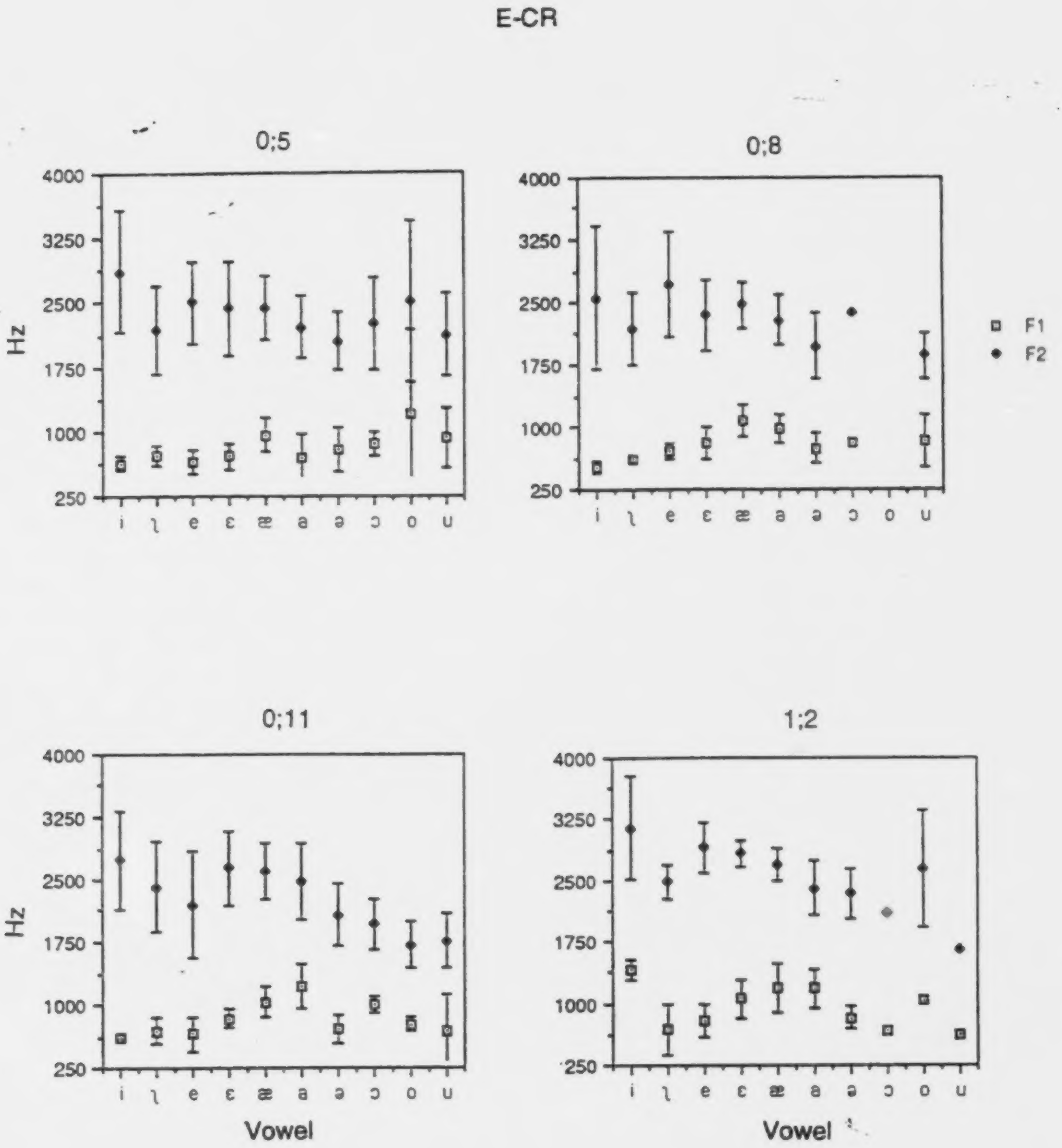


Figure 4. **Mean F1 and F2 and standard deviations for 10 transcribed vowels ([ø] was not included) for the American (E-CR) infant at 0;5, 0;8, 0;11, and 1;2.**

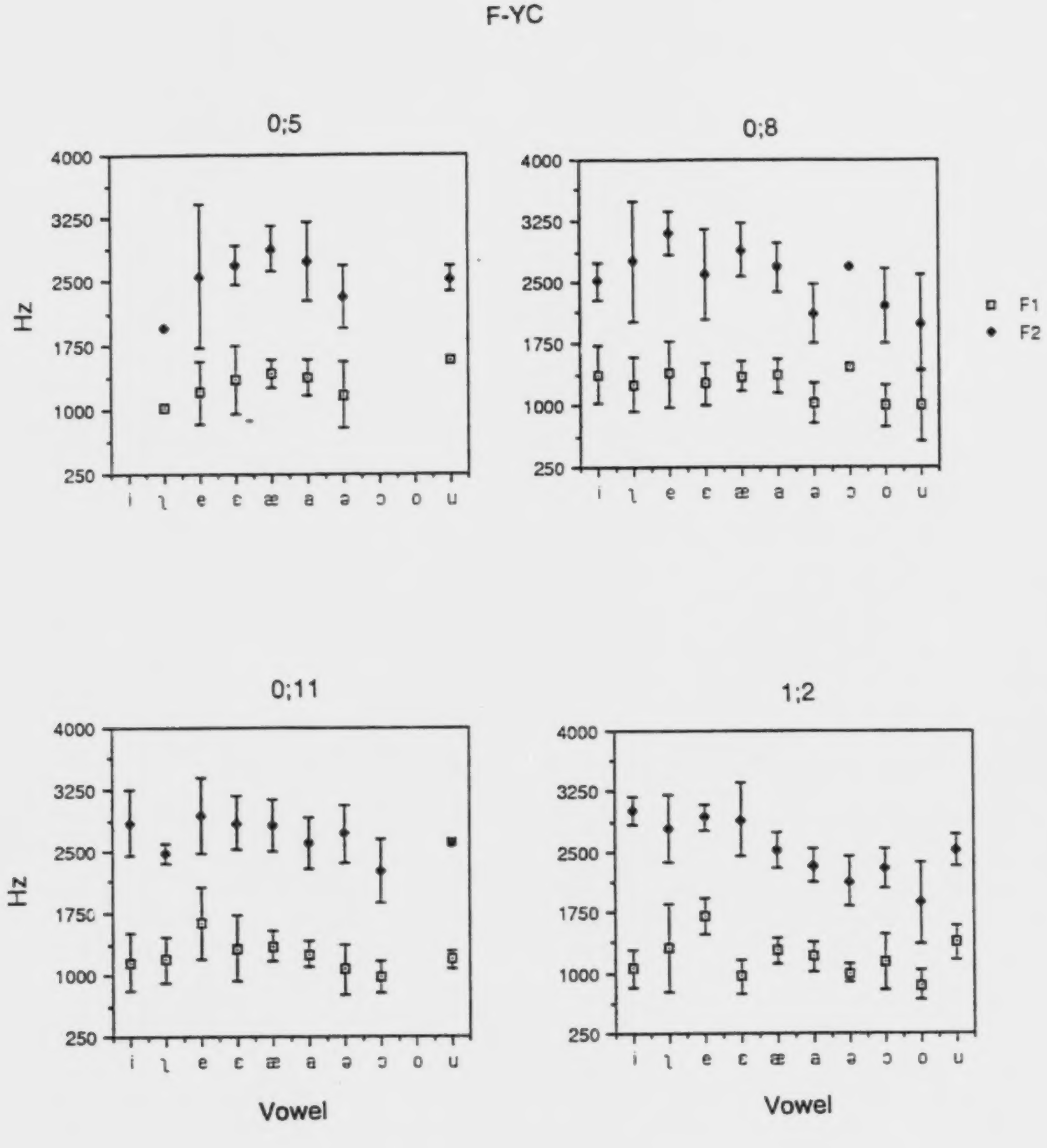


Figure 5. **Mean F1 and F2 and standard deviations for 10 transcribed vowels ([ø] was not included) for the French (F-YC) infant at 0;5, 0;8, 0;11, and 1;2.**

Given the appropriate patterning of F1 in Figure 4, and in particular the low variance in F1 at the first three ages, CR, the English infant, may have developed better jaw control first. The increased variance in F1 at 1;2 for CR appears to be associated with a reduction in variance in F2, which suggests that CR may have begun mastering the articulatory patterns for F2 at that time. On the other hand, given the regular patterning of F2 in Figure 5 from 0;11 months on, YC may have mastered tongue fronting first. This would be reasonable, given the importance of tongue fronting in French (Delattre, 1965). Furthermore, there is a notable reduction in the variance in F1 and/or F2 in YC's productions from 0;11 to 1;2 in the four vowels that are most frequent in French (/e, a, i, ə/).

In order to test whether F-YC produced less variable F2's and E-CR less variable F1's, we conducted an analysis of variance on the log transforms of the standard error of the means for each vowel's F1 and F2 measurements at each time period. The analysis contained one between factor (infant language source), with each of the ten vowels treated as subjects, and two within factors (time period and formant number). There was a significant effect of formant number [$F(1, 18)=27.525$, $p = 0.0001$] with F1 means overall less variable than F2 means. There was also a significant interaction of formant number by language source [$F(1,18)=9.099$, $p = 0.0074$], which showed that F1 measures for E-CR were less variable than those for F-YC, whereas F2 measures for F-YC were less variable than those for E-CR. There were no other significant effects or interactions.

Thus, the analyses of the formant patterns suggest that both infants appeared to develop more adult-like vowel formant frequency patterns over time, presumably as a result of their increasing articulatory control in producing vowels. There also is some evidence for the influence of language-specific articulatory effects in the pattern of variability for F1 and F2 , which reflects the particular ways in which each infant began to demonstrate increasing articulatory control.

Syllable characteristics. All of the two-syllable or longer utterances that contained at least two consonant- and vowel-like segments (e.g., CVCV or VCVC) were tallied for each infant at each age. Figure 6 provides some support for the stages of Stark (1980) and Oller (1980) in terms of the onset of reduplicative babbling (for CR) and the period during which variegated babbling becomes more frequent (for both infants). Thus, the American infant, CR, produced reduplicative or canonical babbling at 0;8 but not before. Both Stark and Oller suggest that such babbling should begin between 0;6 and 0;7, so CR's performance is consistent with this stage in their developmental schemes. On the other hand, the French infant, YC, actually produced some reduplicative utterances in the samples at 0;5, but also showed a great increase in production of such babbles between 0;5 and 0;8. The speech samples of both children yielded a fairly level number of reduplicative babbles from 0;8 on, with the French infant producing a consistently higher number.

Stark and Oller predict that variegated babbling should begin between 0;10 and 0;11 months. Both infants produced variegated babbles before the 11-month sample, but they both showed the greatest number of variegated babbles at 0;11. The fact that both reduplicative (in the case of YC) and variegated babbling (in the cases of both infants) appeared earlier than expected is consistent with the report of Smith, Brown-Sweeney, and Stoel-Gammon (1989), who also found that both reduplicative and variegated productions co-occurred in the babbling of the ten American infants that they studied at four month intervals from 0;6 to 1;6 months of age. (See also Mitchell & Kent, 1990, for similar observations).

We conducted an analysis of variance on the number of reduplicated and variegated babbles with one between factor (infant language source) and one within factor (babbling type - reduplicated or variegated), treating each of the four time periods as subjects. There was a significant effect of infant language source [$F(1,6)=7.670$, $p = 0.0324$], reflecting the higher incidence of multisyllabic babbling in the French infant's speech. There was also a significant effect of babbling type [$F(1,6) = 10.826$, $p = 0.0176$], since there were more variegated than reduplicated utterances overall. There was no infant language source by babbling type interaction. Thus, although the French infant produced more multisyllabic babbles, the pattern of variegated and reduplicative babbling was essentially the same for both children.

The transcription data were also used to determine the number of open and closed syllables produced by both infants at each age. The percentage of closed syllables in the productions of the American infant (CR) dramatically increased at 0;11, while percentages for the French infant remained low and relatively stable (see Figure 7). A chi-squared test on the frequency distributions of open and closed syllables was significant in the case of E-CR (X^2 (3)=33.648, $p = 0.0001$) but not in the case of F-YC (X^2 (3)=5.983, p=0.1124). This discrepancy is consistent with the greater frequency of closed syllables in English than in French (Delattre, 1965) and suggests that by 0;11 CR had begun to produce a distribution of syllable types that is characteristic of his language environment.

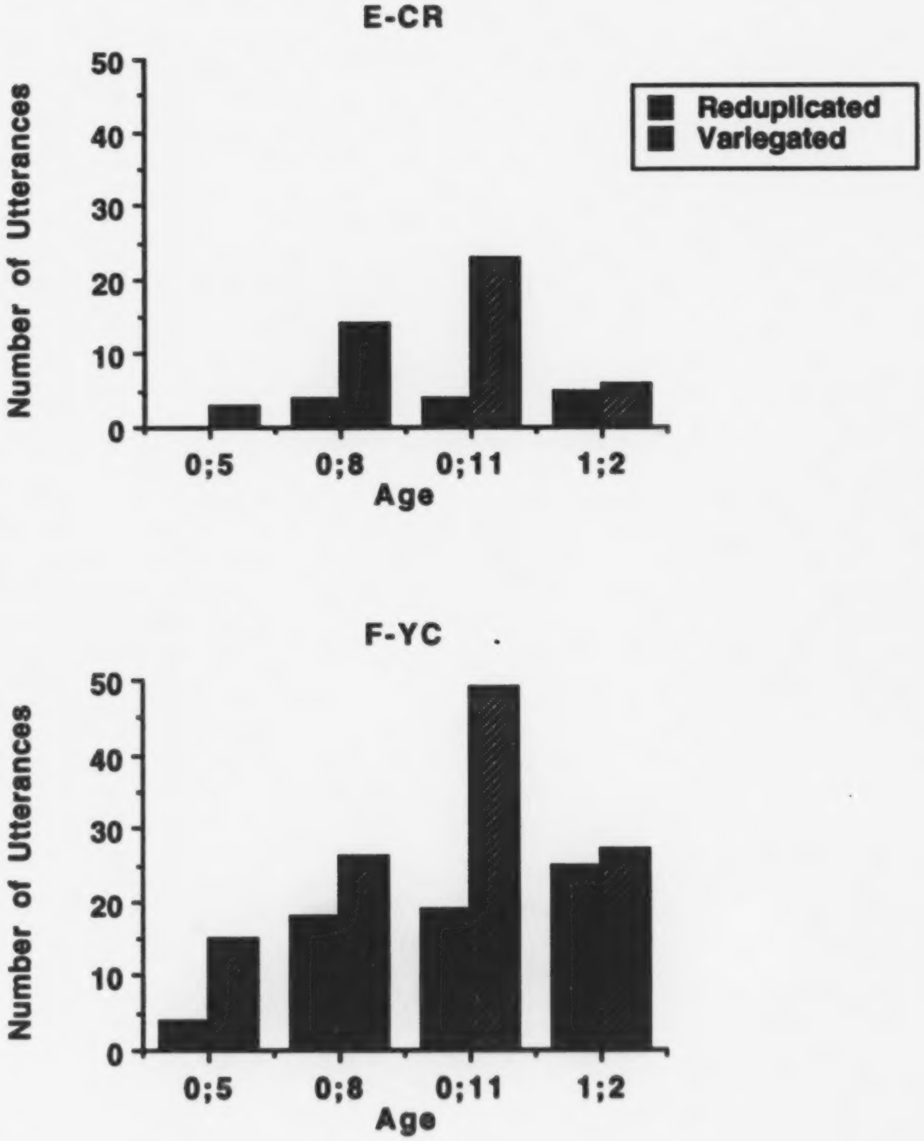


Figure 6. Number of reduplicated and variegated utterances produced by the English (E-CR) and the French (F-YC) infants at 0;5, 0;8, 0;11, and 1;2. The English infant (E-CR) did not produce any reduplicated utterances at 0;5.

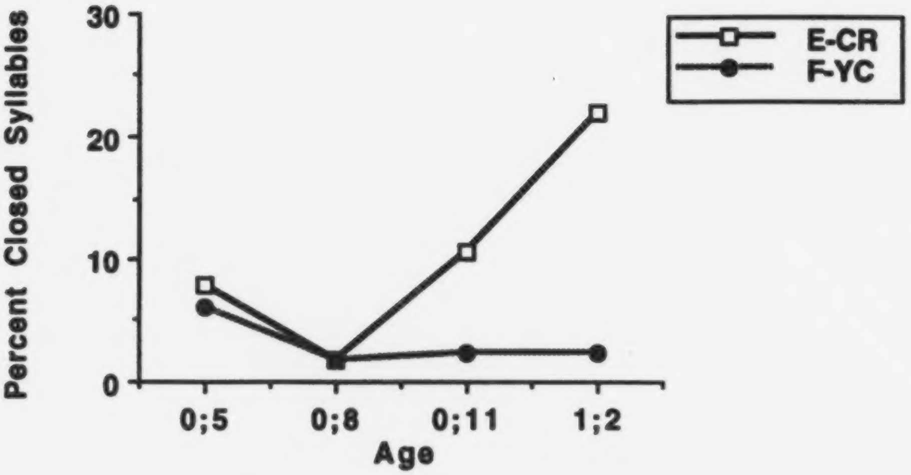


Figure 7. **Percentage of closed syllables produced by the English (E-CR) and the French (F-YC) infant at 0;5, 0;8, 0;11, and 1;2.**

Finally, the number of syllables per utterance for all utterances containing at least one consonant or consonant-like segment was plotted for each infant at each of the four ages. If the children's productions of monosyllabic versus polysyllabic utterances were influenced by the ambient language, then we would expect a greater incidence of polysyllabic utterances produced by the French child and a greater incidence of monosyllabic utterances produced by the American child. X^2 tests of the frequency distributions of the monosyllabic and polysyllabic utterances of the two children at 0;5 and 1;2 were not significant (X^2 (1)=.4, p=0.5251 and X^2 (1)=.1, p = 0.7815, respectively), although they were significant at 0;8 (X^2 (1)= 31.5, *p* = 0.0001) and at 0;11 X^2 (1)= 12.8, (*p* = 0.0003). However, as can be seen in Figure 8, at 0;8 it was actually the American child who produced more polysyllabic utterances, while the French child produced more monosyllabic utterances. At 0;11 the pattern was somewhat more as predicted, in that the French infant produced considerably more polysyllabic than monosyllabic utterances, although the American infant did not produce more monosyllabic than polysyllabic utterances at that time. Since other researchers have found an effect of adult-language utterance lengths on child productions (Boysson-Bardies, 1989) in examining the productions of children beginning at 0;10 and following them until the 25 word stage, it is possible that utterance length effects emerge later in children's productions.

Timing. In terms of rhythmic or timing constraints, French shows greater isosyllabicity than English, or a tendency for non-final syllables to be more or less equal in length. Furthermore, syllable lengthening is also more consistent in French, because it is always breath-group final and is not confounded with lengthening due to lexical stress, as in English. To determine the relative isosyllabicity of syllable duration for the French infant as compared with that of the American infant, the means of all non-final syllables in the first, second, or third position of an utterance, at each of the four ages, were calculated for both F-YC and E-CR. Means for F-YC were generally less variable across utterance position than those for E-CR (see Figure 9).

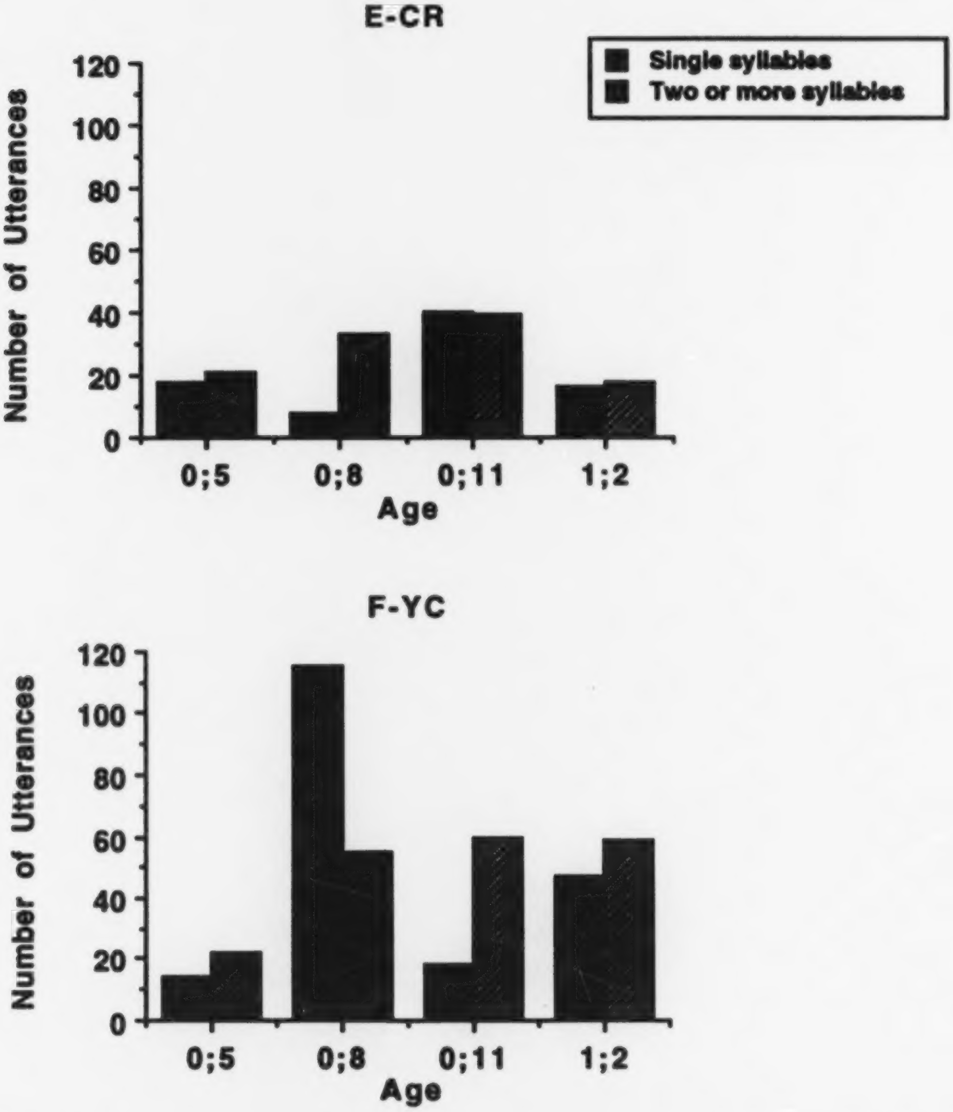


Figure 8. Number of monosyllabic and polysyllabic utterances produced by the English (E-CR) and the French (F-YC) infant at 0;5, 0;8, 0;11, and 1;2.

The difference between the longest average syllable length and the shortest was 62.3, 23.6, 48.7, and 131 ms at 0;5, 0;8, 0;11, and 1;2 respectively, for the French infant YC, whereas for the American infant, CR, this difference was 103.3, 155.8, 80.3 and 144.4 ms at 0;5, 0;8, 0;11, and 1;2. These differences just failed to reach significance, (t(3)=2.1, p=0.0659, one-tailed). The tendency for the French infant to produce more regularly-timed nonfinal syllables appears nonetheless to be a strong one, because at every age, the French infant produced nonfinal syllables that were on average closer in duration to one another than did the American infant. Konopczynski (1986), who examined the duration of syllables in the babbling productions of four French children aged 0;8 to 2;0 months also found evidence for isosyllabicity.

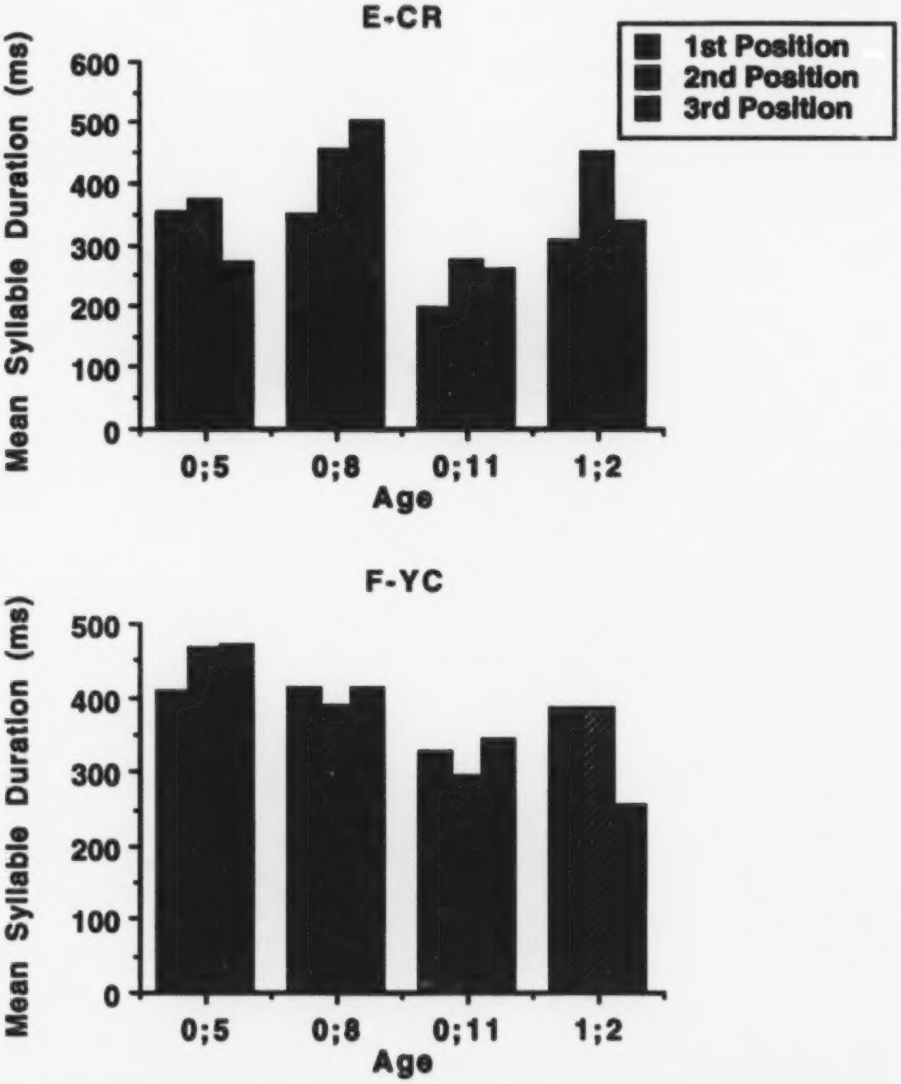


Figure 9. Mean syllable duration in ms for syllables in the first, second, or third position in the utterances produced by the English (E-CR) and the French (F-YC) infant at 0;5, 0;8, 0;11, and 1;2.

Final syllable lengthening in the present study was examined in terms of mean syllable length for final and non-final syllables in reduplicative strings, and in terms of the number of reduplicative strings which showed final syllable lengthening. Reduplicative babbling was used in order to eliminate differences in syllable lengths due to the effects of segments of intrinsically different lengths. Final syllables for both infants were found to be longer than non-final syllables at 0;11 and 1;2 (see Figure 10).

However, when the percentage of reduplicative tokens which have longer final than penultimate syllables is plotted for each infant at each age, the French infant showed a gradual increase in the percentage of utterances with final syllable lengthening, while the American infant showed a drop at 1;2 (see Figure 11). The difference in the number of utterances with final-syllable lengthening for the two children at each of the four ages was significant ($t(3) = 2.377$, $p = 0.049$, one-tailed). Regular final syllable lengthening is a more striking characteristic of the French language, where the final syllable is the only one lengthened, than English, where syllables that bear stress, as well as those that are final, can be lengthened.

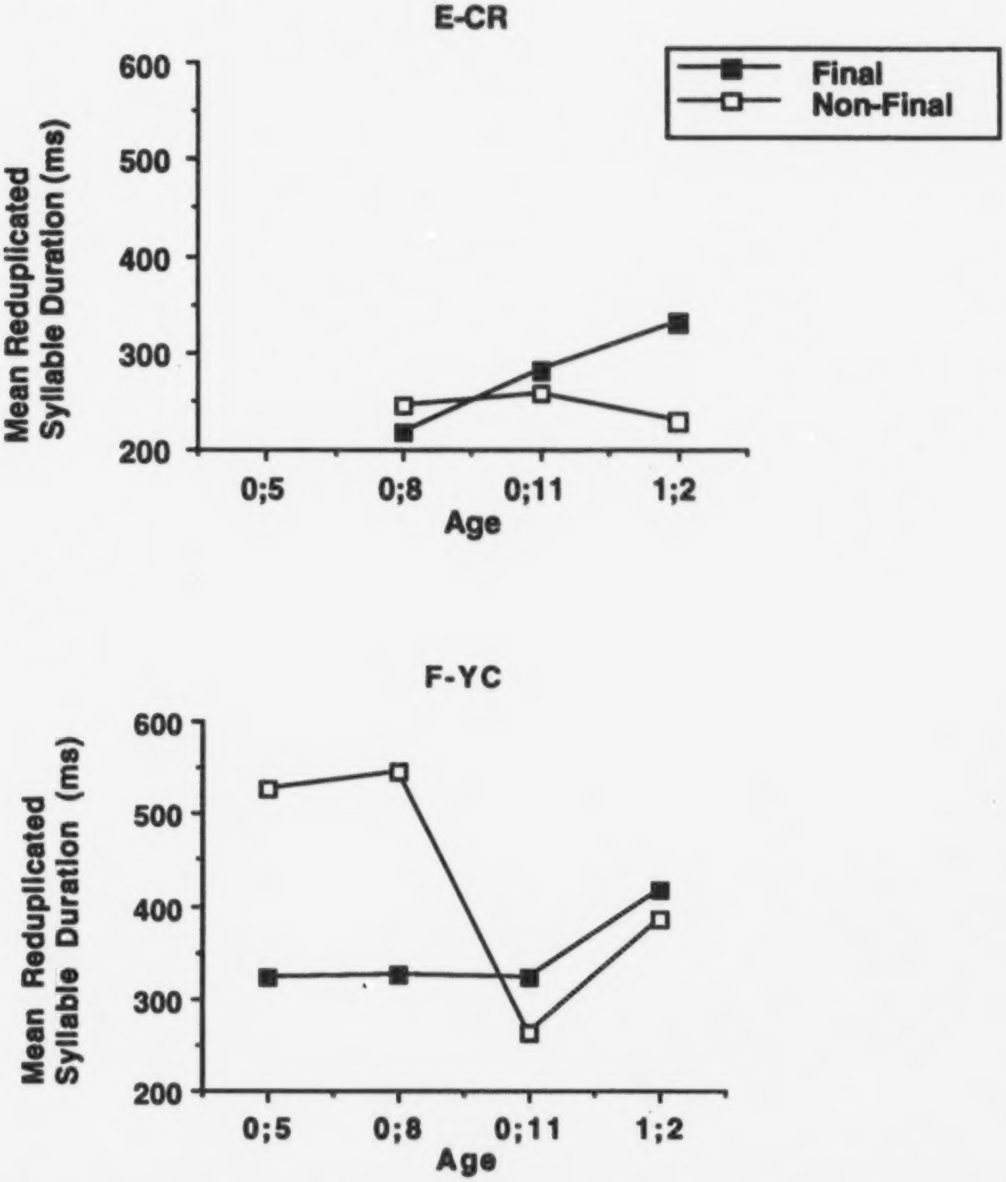


Figure 10. **Mean syllable duration in ms of final versus nonfinal syllables of the reduplicated utterances produced by the English (E-CR) and the French (F-YC) infant at 0;5, 0;8, 0;11, and 1;2. The English infant (E-CR) did not produce any reduplicated utterances at 0;5.**

These data on final syllable lengthening are consistent Konopczynski's (1986) finding of final syllable lengthening in the variegated and reduplicative babbling of French children between the ages of 0;8 and 2;0 months and with Oller and Smith's (1977) finding of little final syllable lengthening in the reduplicative babbling of six American infants aged 0;8 to 1;0. However, as is evident from Figure 6, the American infant produced very few reduplicative utterances overall, so that the present results, while consistent with previous research, must be interpreted with caution.

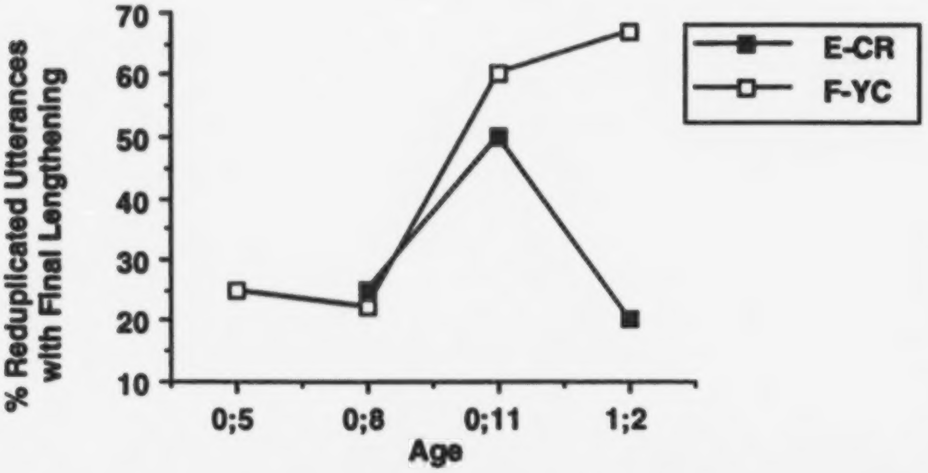


Figure 11. **Percentage of reduplicated utterances with final syllable lengthening produced by the English (E-CR) and the French (F-YC) infant at 0;5, 0;8, 0;11, and 1;2. The English infant (E-CR) did not produce any reduplicated utterances at 0;5.**

GENERAL DISCUSSION

Analyses of the babbling of the two infants in this study showed clear evidence both of similar developmental patterns, presumably due to universal developmental processes in the two infants, and of the prelinguistic influence of the adult language environment.

Evidence of general developmental effects is apparent in both the acoustic analyses and in the examination of segmental inventories. Both infants appeared to produce target F1 and F2 with less variability as they got older, suggesting increased articulatory control. Both infants produced 16 of the 23 consonant-like sounds and all 11 vowel sounds which compose the total corpus, and they generally favored stop consonants over other types. These findings are consistent with Locke's (1983) hypothesis of universal phonological development. Finally, the analysis of the multisyllabic babbling of both children provided some evidence in support of the developmental stages of babbling as described by both Stark (1980) and Oller (1980), although the presence of reduplicative and variegated babbling prior to their predicted onsets is consistent with the findings of Smith, Brown-Sweeney & Stoel-Gammon (1989) and Mitchell and Kent (1990).

The evidence of language-specific effects is also clear, although in most cases these effects appear in longitudinal analyses at the 11-month sample or later for both infants. Each infant's phonetic inventory began to resemble that of the adult language both in composition and frequency, so that over the course of time, the two infants' phonetic inventories, in particular their consonantal inventories, diverged. Vihman, Ferguson, and Elbert (1986) found that the phonetic tendencies of 10 American infants aged 0;9 to 1;4 tended to converge over the period studied. Both the Vihman et al. finding of increased similarity for infants learning the same language and the divergence found in the present analysis for infants learning two different languages suggest the influence of the language environment on the infant productions.

Measurements of the acoustic properties of the vowels also revealed some language-specific effects. Mean F1 for all syllables for both infants appeared to converge over time, whereas mean F2 for the American infant, CR, displayed an increase of about 200 Hz at 1;2, corresponding with a drop of about 300 Hz in mean F2 for the French infant, YC, at that time, so that mean F2 is higher for CR than for YC at 1;2. These F1 and F2 patterns are consistent with the results of a similar analysis by Boysson-Bardies, Halle, Sagart, and Durand (1989), who found that mean F1 measurements were essentially the same for English-learning and French-learning infants, whereas mean F2 measurements were higher for the former group. Furthermore, the patterns found for F1 and F2 in the infant groups reflected similar F1/F2 relationships in the data from the adult languages.

Although this pattern was evident for the babbling of ten-month-old infants in the Boysson-Bardies et al. study, it only emerged in the 14-month sample in our data.

In terms of syllable characteristics, the American infant produced a larger percentage of closed syllables than the French at 0;11 and 1;2, and showed a dramatic increase longitudinally in the percentage of closed syllables produced, while percentages of closed syllables for the French infant remained stable over time. Because English has more closed syllables than French, this result suggests the influence of English on CR's babbling.

Analyses of timing patterns in the babbling of the two infants revealed evidence of the characteristic isosyllabicity of French for F-YC at all ages: when the durations of nonfinal syllables were averaged for each position in the utterance at each age, mean syllable duration tended to be less variable for the French infant than for the American infant. In addition, although both infants showed some final syllable lengthening at 0;11 and 1;2, a larger percentage of the French infant's reduplicative utterances showed final syllable lengthening, suggesting a more pronounced influence of the French rhythmic pattern on F-YC's babbling.

In general, our results agree with recent findings (Vihman et al., 1986; Boysson-Bardies, 1989; Boysson-Bardies et al., 1989) of language-specific influences on the segmental characteristics of babbling towards the end of the first year. However, some of the timing differences that emerged in this study, particularly the suggestion of an early development of isosyllabicity in the babbling of the French infant, suggest that the language environment may affect the prosody of babbling even earlier than its segmental characteristics (see also Whalen, Levitt, & Wang, 1991). Clearly the prosody of infant babbling warrants further cross-linguistic research.

REFERENCES

Atkinson, K. B., MacWhinney, B., & Stoel, C. (1970). An experiment in the recognition of babbling. *Papers and Reports in Child Language Development, 1,* 71-76.

Bond, Z. S., Petrosino, L., & Dean, C. R. (1982). The emergence of vowels: 17 to 26 months. *Journal of Phonetics, 10,* 417-422.

Boysson-Bardies, B. de (1989) Material evidence of infant selection from the target language: a cross-linguistic phonetic study. Paper presented at the Conference on Phonological Development, Stanford University, September, 1989.

Boysson-Bardies, B. de, Halle, P., Sagart, L., & Durand, C. (1989). A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language, 16,* 1-17.

Boysson-Bardies, B. de, Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language, 11,* 1-15.

Buhr, R. (1980). The emergence of vowels in an infant. *Journal of Speech and Hearing Research, 23,* 73-94.

Brown, R. (1958). *Words and things.* New York: Free Press.

Delattre, P. (1965). *Comparing the phonetic features of English, German, Spanish and French.* Heidelberg: Julius Groos.

Dinger, M. C., & Blom, J. G. (1973). An investigation of infant babbling. *Proceedings from the Institute of Phonetic Sciences* (University of Amsterdam), 3, 42-50.

Ferguson, C. (1978). Learning to pronounce: The earliest stages of phonological development in the child. In F.D. Minifie & L. L. Lloyd (Eds.), *Communication and cognitive abilities—Early behavioral assessment.* Baltimore: University Park Press.

Jakobson, R. (1968). *Child language, aphasia, and phonological universals* (A. Keiler, Trans.). The Hague: Mouton.

Kent, R. D., & Bauer, H. R. (1985). Vocalizations of one-year-olds. *Journal of Child Language, 12,* 491-526.

Konopczynski, G. (1986). Vers un modèle développemental du rhythme Français: Problèmes d'isochronie réconsidérés à la lumière des données de l'acquisition du langage. *Bulletin de l'Institut de Phonétique de Grenoble, 15,* 157-190.

Ladefoged, P. (1975). *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Locke, J. (1983). *Phonological acquisition and change.* New York: Academic Press.

Mack, M., & Lieberman, P. (1985). Acoustic analysis of words produced by a child from 46 to 149 weeks. *Journal of Child Language, 12,* 527-550.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7,* 41-75.

Menn, L. (1971). Phonotactic rules in beginning speech. *Lingua, 26,* 225-241.

Menyuk, P. (1971). The acquisition and development of language. Englewood Cliffs, NJ: Prentice-Hall.

Menyuk, P., & Menn, L. (1979). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), *Language acquisition.* New York: Cambridge University Press.

Mitchell, P. R., & Kent, R. D. (1990). Phonetic variation in multisyllabic babbling. *Journal of Child Language, 17,* 147-165.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology, Vol. 1: Production.* New York: Academic Press.

Oller, D. K., & Smith, L. (1977). Effect of final-syllable position of vowel duration in infant babbling. *Journal of the Acoustical Society of America, 62,* 994-997.

Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language, 3,* 1-11.

Olney, R., & Scholnick, E. (1976). Adult judgments of age and linguistic differences in infant vocalizations. *Journal of Child Language, 3,* 145-150.

Pike, K. (1945). *Intonation of American English.* Ann Arbor, MI: University of Michigan Press.

Smith, B. L., Brown-Sweeney, S., & Stoel-Gammon, C. (1989). A quantitative analysis of reduplicated and variegated babbling. *First Language, 9,* 175-190.

Stark, R. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology, Vol. 1: Production.* New York: Academic Press.

Thevenin, D. M., Ellers, R. E., Oller, D. K., & Lavoie, L. (1985). Where's the drift in babbling drift. A cross-linguistic study. *Applied Psycholinguistics, 6,* 3-15.

Vihman, M., Macken, M., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language, 61,* 397-443.

Vihman, M., Ferguson, C., & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics, 7,* 3-40.

Weir, R. (1966). Some questions on the child's learning of phonology. In F. Smith & G. A. Miller (Eds.), *The genesis of language.* Cambridge, MA: MIT Press.

Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10,* 193-216.

Whalen, D. H., Levitt, A. G. & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language, 18,* 501-516.

FOOTNOTES

[*] *Journal of Child Language,* in press.

[†] Also Wellesley College.

[††] Wellesley College. Now at Brown University.

[1] The two transcribers of the vocalizations used in this analysis are both native speakers of English who are fluently bilingual in French. However, it is possible that when the data for the French infant were considered, the transcriptions of ambiguous sounds, particularly vowels, might be interpreted in favor of English targets. For example, the vowel sound of 'beet,' when short in duration, might be interpreted as the vowel sound of 'bit,' a frequent sound in English, but one which does not exist in French. Of course, any such bias would lead to conservative estimates of language-specific differences between the infants.

[2] The data at 1;2 may have included some words with the babble, because it was not possible to identify and exclude words on the basis of the parents' written comments during that month. However, we observe that the pattern here, and others described later, in general begin at 0;11 or earlier, and infants were clearly prelinguistic then. Thus, we do not feel that the possible inclusion of some words in the data at 1;2 substantially changes our pattern of results.

[3] The mean formant frequencies for English in Boysson-Bardies et al. (1989) were calculated for British English. We are assuming that the values for American English would bear essentially the same relationship to the other languages as did British English, although the actual means might differ.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 63-80

Listening with Eye and Hand: Cross-modal Contributions to Speech Perception*

Carol A. Fowler† and Dawn J. Dekle††

Three experiments investigated the basis for the "McGurk effect" whereby optically-specified syllables experienced synchronously with acoustically-specified syllables integrate in perception to determine a listener's auditory perceptual experience. One hypothesis is that the effect arises when optical and acoustic cues for a syllable are associated in memory. A second hypothesis is that the effect arises when cross-modal information, familiar or not, is convincingly about the same articulatory speech event in the environment. Experiments contrasted the cross-modal effect of orthographic syllables on acoustic syllables, presumed to be associated in experience and memory, with that of haptically-experienced and acoustic syllables, presumed not to be associated. Findings were that the latter pairing, but not the former, gave rise to cross-modal influences under conditions in which subjects were informed that cross-modal syllables were paired independently. Mouthed syllables felt by a perceiver affected reports of simultaneously heard syllables (and reports of the mouthed syllables were affected by the heard syllable). These effects were absent when syllables were simultaneously seen (spelled) and heard. We conclude that the McGurk effect does not arise due to association in memory, but rather to conjoint near specification of the same causal source in the environment--in speech, the moving vocal tract producing phonetic gestures.

In a variety of circumstances, including, presumably, face-to-face spoken communications outside the laboratory, listeners can acquire phonetic information optically as well as acoustically. Seeing the face of a speaker considerably improves listeners' abilities to recover speech produced in noise (e.g., Erber, 1969; Ewertsen & Nielsen, 1971; Sumby & Pollack, 1954), and, when a visible talker and his or her apparent acoustic output are mismatched in an experiment, as in the so-called "McGurk effect" (e.g., McGurk & MacDonald, 1976; MacDonald & McGurk, 1978), phonetic information recovered optically may override that recovered acoustically. This is particularly likely to occur when the phonetic information is for consonantal places of articulation close to the front of the speaker's mouth. Accordingly, optical "tap" paired with acoustic "map" may be reported as "nap," with voicing and nasality consistent with the acoustic signal and place of articulation consistent with the optical display (MacDonald & McGurk, 1978; Summerfield, 1987).

Remarkably, the cross-modal influence, phenomenally, is not due to hearing one utterance, seeing another and reporting some compromise. Rather the visible utterance generally changes what the listener experiences hearing (Liberman, 1982; Summerfield, 1987). Accordingly, the visual influence remains when subjects are instructed to report specifically what they heard rather than what the speaker said (e.g., Summerfield & MacGrath, 1984), and it remains even after considerable practice attending selectively (Massaro, 1987).

Two general accounts of the McGurk effect may be derived from current theories of speech perception. One is that perceivers consult memory representations of fundamental units of spoken

---

The research reported here was supported by NICHD Grant HD-01994 to Haskins Laboratories. We thank George L. Wolford for his participation in the pilot research, his guidance on some of the statistical analyses and for his comments on an earlier draft of the manuscript; we also thank Lawrence Rosenblum for comments on the manuscript and Michael Turvey for comments on our General Discussion.

utterances (prototypes of syllables in Massaro's theory; e.g., 1987; 1989) that include specifications both of optical and acoustic "cues" for the utterance. Presumably, the memory representations derive from experience with token productions of the utterance in which various subsets of the optical and acoustic cues were detected by the perceiver. When partially conflicting optical and acoustic cues are detected in a McGurk procedure, the listener selects—and experiences hearing—the memory representation most consistent with the collection of cues. In this kind of account, the influence of the visual display on the percept derives from the association of optical and acoustic cues in memory, and these associations, in turn, derive from the association of the cues in the world as sampled by the perceptual systems.

A different general account of the phenomenon derives from theories of speech perception claiming that listeners to speech do not hear the acoustic signal *per se*, but rather hear the phonetically-significant gestures of the vocal tract that give rise to the acoustic speech signal. Two such theories are the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) and the direct-realist theory (Fowler, 1986; Rosenblum, 1987).

The motor theory was developed to explain findings suggesting that, at least in some circumstances, there is a closer correspondence between the listener's percept and the vocal-tract gestures of the talker than between the percept and the acoustic signal. (See Liberman et al., 1967 for some examples.) According to the motor theorists, a speech mode of perception in which articulation is recovered from acoustics evolved to handle the special problems associated with recovering phonetic information encoded in the acoustic signal. The encoding occurs because consecutive consonants and vowels in an utterance are coarticulated—that is, produced in overlapping time frames. In consequence of coarticulation, there are no boundaries *between* phonetic segments in the signal, and the acoustic information for a consonant or vowel is highly context-sensitive. According to the theory, listeners use the acoustic speech signal to devise a hypothesis concerning the set of articulatory gestures that, when coarticulated, would have given rise to that signal. Testing the hypothesis involves the central component of the listener's speech motor system (an "innate vocal tract synthesizer" according to Liberman & Mattingly, 1985) and that, in the theory, is how the percept acquires a motor character.

The direct-realist theory accepts the motor theorists' evidence that listeners to speech recover phonetically-significant gestures of the vocal tract, but it explains the recovery differently. In the theory (derived from Gibson's more general theory of direct perception; e.g., 1966, 1979), perception is the only means by which organisms can know the environment in which they participate as actors. It is crucial, therefore, that perceptual systems generally acquaint perceivers with relevant aspects of the environment. Perceptual systems do so by extracting information from media, such as light, skin and air, about properties of the environment. These media can provide information about the environment, because properties of the environment cause distinctive patternings in them; the distinctive patterns, in turn, can serve as information for their causal source. In vision, although retinas are stimulated by reflected light, perceivers do not see patterns in light; rather, they see the environmental sources of those patterns. In haptics, perceivers do not infer palpated object properties from felt skin deformations; they feel the object sources of the skin deformations themselves. By analogy, in hearing, and specifically in speech perception, listeners should not perceive the acoustic signal, but rather its cause in the environment. In speech, the immediate cause of the signal is the vocal tract activity of the talker.

In either theory, the motor theory or the direct realist theory, the McGurk effect, and audio-visual influences on speech perception more generally, arise because the optical and acoustic information is convincingly about the same speech event, and speech events, not structured media, are perceptual objects.

In the following experiments, we attempt to distinguish the two general accounts we have offered for the McGurk effect. In its global character, the effect is consistent with both accounts. Perceivers frequently both see and hear a speaker, and so they have ample opportunity to develop memory representations including both optical and acoustic cues. By the same token, outside the laboratory, visible talking and audible talking emanating from the same location in space *are* the same event of talking; if the objects of perception are environmental, as specified by information in media, then it is expected that information in different media that are joint consequences of the same event (or, in the

McGurk procedure, of ostensibly the same event) serve jointly to specify the event to the perceiver.

In the following experiments, we have attempted to distinguish the two accounts by looking for cross-modal influences on speech perception from two new sources each of which captures one, but not the other, distinctive aspect of the McGurk paradigm that might account for the cross-modal influences there.

One situation was meant, on the one hand, to capture the association in experience, and hence in memory, of an acoustically-specified utterance with a specification in another modality. On the other hand, it was meant to exclude association via conjoint lawful specification of a common environmental event. To achieve this, we paired spoken syllables with matched or mismatched orthographic representations of syllables (cf. Massaro, Cohen, & Thompson, 1988). Our college-student subjects have been readers of an alphabetic writing system for over a decade, and they experience redundant pairings of sight and sound whenever they read aloud or else see a text that someone else is reading aloud. Although listeners may be less experienced with sound-spelling pairings than with pairings of the sound and sight of a speaker, their experience with the former pairings are sufficient that a spoken monosyllabic word can activate its spelling for a listener. For example, detection of an auditorily-presented word that rhymes with a cue word also presented auditorily is faster if the two words are spelled similarly (e.g., "pie"-"tie") than if they are spelled differently (e.g., "rye"-"tie") (Seidenberg & Tanenhaus, 1979; see also Tanenhaus, Flanigan, & Seidenberg, 1980). (In an experiment similar to the orthographic condition of our Experiment 1, Massaro et al. (1988) reported weak cross-modal influences of a written on a spoken syllable. Our Experiments 1 and 2 explore the conditions under which this occurs and compare the effect to another possible source of cross-modal influences on speech perception.)

An important characteristic of the acoustic-orthographic pairings for our purposes is that their association is by convention, rather than by lawful causation (cf. Campbell, 1989). That is, while optical and acoustic correlates of a speaker talking are associated in the world because they are lawful consequences of the same event of talking, graphemes and their pronunciations are associated in the world by societal convention.

The second experimental situation that we devised was meant, insofar as possible, to be complementary to the first. That is, we established a cross-modal pairing that is unfamiliar to subjects, but that—outside the laboratory—is a lawful pairing, because the same environmental event gives rise to structure in the two different media. In this situation, we paired acoustically-specified syllables with matched and mismatched manually-felt, mouthed syllables. Our guess, confirmed by our subjects, was that they did not recollect experiences in which they had handled someone's face while he or she was talking. Although some of them may have had such experiences—as infants or young children perhaps—they must have had considerably less experience of that sort than they have had either seeing and hearing spoken utterances or seeing and hearing text being read.

We know that some phonetic information can be obtained by feeling the face and neck of a speaker. Helen Keller learned to speak English, French and German by obtaining phonetic information haptically (e.g., Keller, 1903). Other deaf-blind individuals have learned some speech and have learned to understand spoken language remarkably well using the Tadoma method, in which they learn to detect phonetic properties of speech haptically (e.g., Chomsky, 1986; Schultz, Norton, Conway-Fithian, & Reed, 1984). We do not know whether inexperienced perceivers can recover phonetic information by touch, and it is important for our test that our subjects not be trained. Accordingly, we selected a fairly distinct articulatory difference for them to detect, the difference between /ba/ and /ga/.

Our expectations were as follows. If the operative factor in the McGurk effect is the association in memory of cross-modal cues available during events in which speech occurs, then an influence of written syllables on heard syllables should occur, but an effect of felt on heard syllables should be weak or absent. If the operative factor is, instead, the (apparently) common causal source in the environment of the acoustic and optical structure, then felt syllables will affect what listeners report hearing, while orthographic representations of spoken syllables will not. Alternatively, of course, both (or neither) factor may be important.

EXPERIMENT 1

Our first experiment tested these predictions by looking for cross-modal haptic and orthographic influences on identifications of heard syllables and for reverse effects of heard syllables on reports of felt and read syllables. In this experiment and the next, we informed subjects that cross-modal syllables were paired independently, and we requested that, therefore, they make their judgments of heard syllables based only on what they heard; we did so in an effort to restrict cross-modal effects to influences on perception rather than on judgment.

Methods

Subjects. Subjects were 23 undergraduate students at Dartmouth College. All were native speakers of English who reported normal hearing and normal or corrected vision. None of the participants was experienced in the Tadoma method and none had special training in lipreading. Of the 23 participants, one student's data were eliminated from the analyses when she reported in debriefing that she had not believed our (accurate) statement that felt and heard syllables had been paired independently. Data from 10 additional subjects were eliminated from most analyses, because their identifications of felt syllables did not significantly exceed chance (60.6% correct; $z = 1.64$).

Stimulus materials. Acoustic stimuli were three-formant synthetic consonant-vowel (CV) syllables produced by the serial resonance synthesizer at Haskins Laboratories. There were 10 syllables in all, ranging across an acoustic continuum of F2 transitions from a rising transition appropriate for /ba/ to a falling transition appropriate for /ga/. F3 was fixed and rising across the continuum; accordingly, there were no intermediate /da/ syllables. Steady-state values for the three formants were 800, 1190 and 2400 Hz (with bandwidths of 50, 100 and 200 Hz, respectively). Starting frequencies of F1 and F3 were 500 and 2100 Hz. Starting frequencies of F2 ranged from 760 Hz at the /b/ end of the continuum to 1660 Hz at the /g/ end in 100 Hz steps. Transitions were 50 ms in duration with a following 150 ms steady-state. Fundamental frequency increased by 10 Hz over the first 30 ms of each syllable to a steady-state value of 120 Hz and declined by 10 Hz over the last 40 ms. The amplitude contour had an analogous shape and time course.

The ten continuum members were stored (filtered at 10 KHz and sampled at 20 KHz) in a small computer (New England Digital Company) programmed to run the various conditions of the experiment. They were presented to listeners over a loudspeaker situated to the right of the CRT screen.

In the orthographic condition, on each trial, "BA" or "GA" was printed in upper case on the CRT screen simultaneously with the presentation of the acoustic syllables; the printed syllable remained on the screen until subjects pressed a key to initiate the next trial.

In the Tadoma condition, mouthed syllables were /ba/ and /ga/ produced by a model (the first author). The model faced a CRT screen that specified the syllable to be mouthed on each trial and provided a countdown permitting the mouthed syllables to be produced in synchrony or near synchrony with the acoustically presented syllable.

A single, 60-item test order of the acoustic continuum members was used for all conditions of the experiment. Six tokens of each continuum member were presented in random order but with the constraint that each synthetic syllable occur twice in each third of the test order. In the orthographic condition, a second 60-item test order determined which orthographic syllable would be paired with each acoustic syllable. The same test order dictated the order of syllables to be mouthed in the Tadoma condition. This sequence was also random, but now with the constraint that each orthographic (felt) syllable be paired with each synthetic syllable once in each third of the test order.

Procedure. Each subject participated in three tests, synthetic syllables alone (auditory condition), synthetic syllables paired with printed syllables (orthographic) and synthetic syllables paired with felt syllables (Tadoma). The order of the conditions was counterbalanced, with two subjects (of the twelve who exceeded chance in identifying felt syllables) experiencing each order.

Subjects were run individually. They first heard the endpoint syllables from the synthetic-speech continuum. The endpoints were identified for them and played several times. Subjects were told that the speech was produced by a computer and that they would be identifying syllables like the ones they had just heard in subsequent tests.

In the auditory test, subjects were seated in front of the CRT screen. Printed on the screen was the message "PRESS RETURN TO PROCEED." To initiate each trial, subjects pressed the return key. On each trial, one synthetic syllable was presented over the loudspeaker. Subjects made their responses by circling printed "B" or "G" on an answer sheet; they were instructed to guess if necessary, and then to continue the test at their own pace by pressing the return key for each trial.

In the orthographic condition, the test was similar except that when the return key was pressed, a printed syllable appeared on the screen with its onset simultaneous with the onset of the synthetic syllable. The printed syllable remained on the screen until the subject pressed the return key for the next trial.[1] Subjects were instructed to watch the screen as the printed syllable was displayed. They then made two responses, first circling either "B" or "G" under the heading "heard," indicating which syllable they had heard and then circling "B" or "G" under the heading "saw" indicating which syllable they had seen. Subjects were told explicitly that the acoustic and spelled syllables were independently paired so that they should always base their "heard" judgment on what they had heard independently of what they had seen and vice versa for the "saw" judgment. As in the auditory condition, they were instructed to guess if they were unsure of the syllable they had heard or seen on any trial.

In the Tadoma condition, the model stood facing the CRT screen with the loudspeaker directly in front of her at about waist level. She sequenced the trials by pressing the return key to initiate each one. In advance of presentation of the acoustic syllable, the computer printed a syllable on the screen that the model was to mouth. Then it presented a countdown consisting of a sequence of two asterisks and then a pair of exclamation points (i.e., *...*...!!) at 1000 ms intervals. The exclamation points were presented simultaneously with the onset of the acoustic syllable. Between trials, the model kept her lips parted and jaw slightly lowered. With practice, she learned to time her closing for the mouthed consonant so that the opening coincided phenomenally with onset of the acoustic signal.

Subjects received no instructions at all on how to distinguish felt "ba" from "ga." Each subject stood back to the CRT screen, approximately a step farther from the CRT screen than the model. (This was to prevent the subject from looking at the model's face; in any case, relevant parts of her face were covered by the subject's hand.) The subject stood with his or her right hand in a disposable glove placed over the lips of the model. After some piloting, we found that subjects were most successful if we placed the the forefinger on the upper lip and the next finger on the lower lip of the model. This is not the procedure used by Tadoma perceivers (who generally place a thumb vertically against the lips); however, it allowed subjects to distinguish open from closed lips. After presentation of paired felt and heard syllables, subjects indicated to a second experimenter which syllable, "ba" or "ga," they had heard and then which syllable they had felt, in each case guessing if necessary. They made their responses by pointing to printed syllables on an 8 1/2 × 11 inch sheet of paper held on a clipboard by a second experimenter (the second author). On the sheet of paper, the printed syllables "BA" and "GA" appeared twice, on the left under the heading "heard" and on the right under the heading "felt." The second experimenter then marked an answer sheet analogous to those used in the orthographic condition (that is with response columns "heard" and "felt"). As in the orthographic condition, subjects were told explicitly that synthetic and mouthed syllables were paired independently and that they should, therefore, make their "heard" judgment based only on what they had heard and their "felt" judgment based only on what they had felt.

Because the Tadoma procedure was quite taxing, not only for the model, but also for subjects (whose right arms were extended to the side at the model's mouth level), subjects were allowed to stop and rest at any point during the procedure. Except for rest periods, trials were sequenced by the model pressing the return key immediately after the second experimenter indicated that she had recorded the subject's responses.

The model was, of course aware of the syllable to be mouthed, but not of the synthetic syllable continuum member to be presented on each trial. The experimenter who recorded the subject's responses was blind to the syllable being mouthed, but able to hear the synthetic syllable. Accordingly, neither experimenter had information that could bias their performance of their respective roles in the experiment in respect to predictions concerning effects of felt on heard speech.

Results

Results in the auditory condition for the 12 subjects whose performance identifying "felt" syllables was significantly better than chance are presented in Figure 1. Performance averaged 90 percent "b" judgments for the first two continuum members and 12 percent "b" judgments over the last two. In the cross-modal tasks, performance identifying the orthographic syllables was, not surprisingly, near 100%, while performance judging felt syllables averaged 78.6%.

Effects of the orthographic and felt syllables on "heard" judgments are presented in the top and bottom panels of Figure 2. The top panel shows a small effect of the orthographic syllable on percent "b" judgments, while the bottom panel shows a larger, more consistent effect of the felt syllables. In an analysis of variance with repeated measures factors, Continuum number, Cross-modal syllable (ba/ga) and Condition (Orthographic, Tadoma), the effect of Continuum number on percent "b" judgments was, of course, highly significant ($F(9,99) = 67.52$, $p < .001$), accounting for 56.2% of the variance in the analysis. The effect of Condition was not significant, but the effect of Cross-modal syllable reached significance as did its interaction with condition ($F(1,11) = 17.22$, $p = .0017$; $F(9,99) = 4.85$, $p = .048$, respectively).[2] To explore the basis for the interaction, we performed tests separately on the data from each modality, setting alpha now at .025. The 17.5% difference in heard "b" responses accompanied by felt "ba" versus "ga" was highly significant ($F(1,11) = 13.70$, $p < .001$) while the analogous 5.8% effect in the orthographic condition was marginal ($F(1,11) = 4.52$, $p = .055$). In the original analysis, no other interactions reached significance. The main effect of cross-modal syllable and its interaction with condition remained significant with nearly identical F and p values when the analyses were redone on percent "b" responses averaged across the ten continuum members.

In the next analysis, we looked only at those "heard" trials on which subjects had correctly identified the "felt" syllables. Because this excluded many trials, we examined performance averaged across the continuum. On average, listeners showed a 30.4% difference in favor of heard "b" judgments for those trials accompanied by felt "ba" rather than felt "ga." This difference is nearly double that found by looking at all heard trials, and it is highly significant ($t(11) = 4.25$, $p = .0014$). We did not perform an analogous test on the orthographic data, because performance identifying printed syllables was near 100%.

A further analysis was performed on data from the Tadoma condition, now with data from the auditory condition included. If felt "ba" increases heard "b" judgments, and felt "ga" decreases them, then the means for the auditory condition should fall between those for the felt "ba" and "ga" trials.

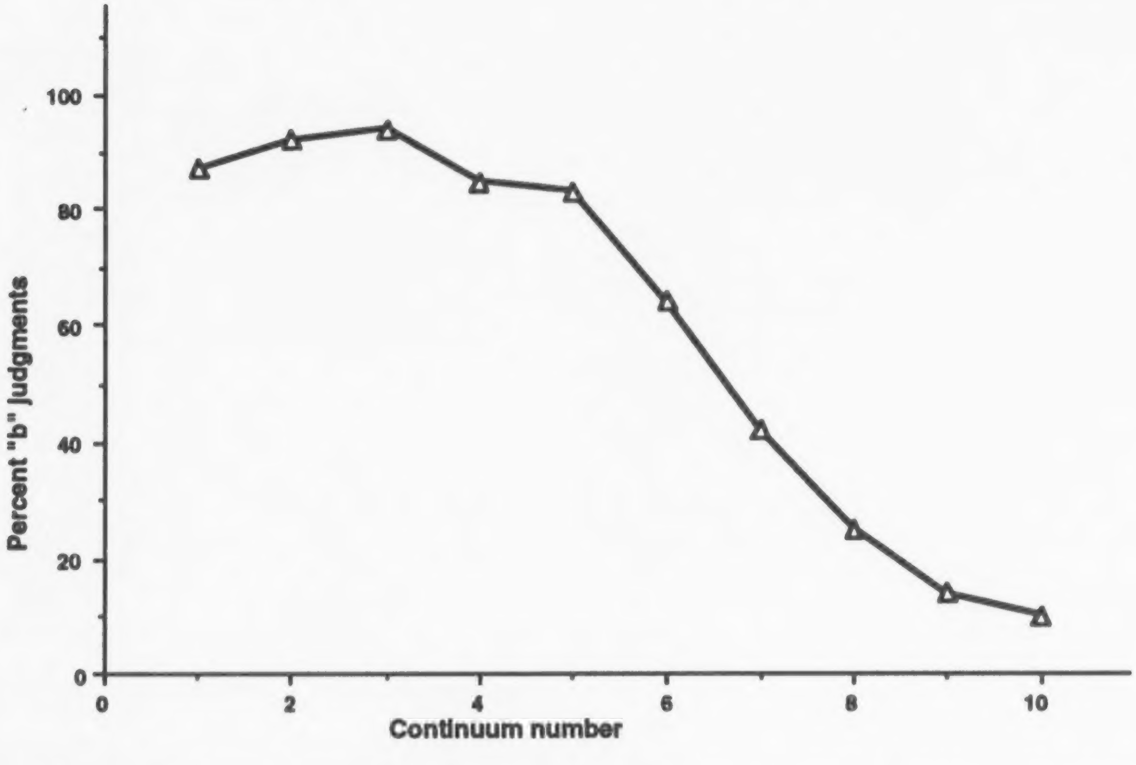


Figure 1. Percent "b" judgments to acoustic continuum members in the auditory condition, in which acoustic syllables were unaccompanied by spelled or felt syllables.

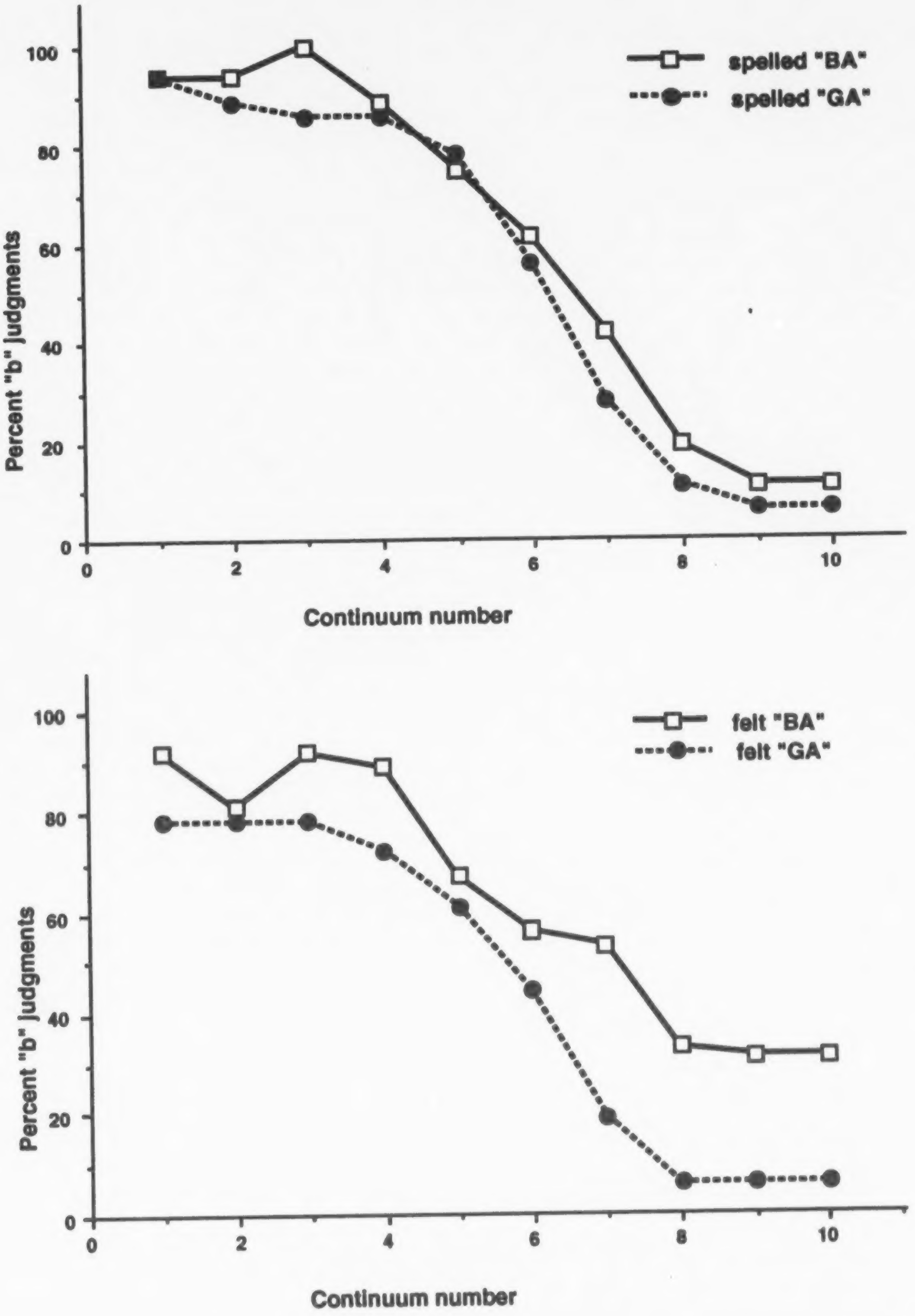


Figure 2. **Percent "b" judgments to acoustic continuum members in cross-modal conditions of Experiment 1. Top panel: orthographic condition; bottom panel: Tadoma condition.**

On average, they do, with percent "b" judgments averaging 62.2% overall on felt "ba" trials, 59.6% on acoustic-alone trials and 44.7% on felt "ga" trials. However, while the difference between the acoustic and felt "ga" trials is large and consistent, that between the acoustic and felt "ba" trials is not. In an analysis of variance with factors Continuum number and Condition (felt "ba," auditory, felt "ga"), both main effects were significant. However, in analyses performed on each comparison separately, only the auditory-"ga" difference was significant ($F(1,11) = 22.66$, $p = .0006$); the auditory-"ba" difference did not approach significance ($F<1$). However, the auditory-"ba" variable did interact with Continuum number ($F(9,99) = 2.09$, $p = .037$), with the predicted direction of effect occurring just on numbers 1,4 and 7-10. The failure of felt "ba" to differ from the auditory condition may signify that only felt "ga" gave rise to a McGurk-like effect. An alternative interpretation that we address in Experiment 3 is that blocking the acoustic-alone trials from the Tadoma trials led to a criterion difference in classification judgments across the conditions.

In another analysis on heard syllables in the Tadoma condition, we examined performance on the first block of 20 Tadoma trials to see whether a cross-modal influence was present from the very beginning, when subjects were least experienced with the task. This test is of particular interest in the Tadoma condition, because subjects had essentially no prior experience handling the faces of talkers. We had designed the test order so that every continuum member occurred once with felt "ba" and once with felt "ga" in the first block of twenty trials. This analysis, performed on "b" responses averaged across the continuum, yielded a significant difference between felt "ba" and "ga" trials ($t(11) = 2.84$, $p = .016$). The difference was 17.5% in favor of "b" responses, a difference, coincidentally, of exactly the same magnitude as the difference averaged over all 60 trials.

We looked for a reverse effect of heard syllables on felt judgments. (We did not look for an analogous effect on orthographic judgments, because performance there was at ceiling.) Massaro (1987) obtained an analogous influence of heard syllables on judgments of visually-perceived mouthed syllables. In an analysis of variance with factors Continuum number and Felt syllable (felt "ba" or "ga"), both main effects and the interaction were significant (Continuum: $F(9,99) = 5.61$, $p < .001$; Felt syllable: $F(1,11) = 61.41$, $p < .001$; interaction: $F(9,99) = 2.08$, $p = .038$). The main effect of Continuum was significant, because there were more felt "ba" judgments on trials associated with acoustic syllables at the /b/ end of the acoustic continuum than on trials associated with acoustic syllables at the /g/ end. The main effect of Felt syllable was significant, because there were more felt "ba" judgments when the felt syllable was "ba" (85.9%), than when it was "ga" (29%). To explore the basis for the interaction, we performed separate analyses on the felt "ba" and felt "ga" trials. In both instances, the effect of Continuum was significant; we used trend tests to look for a linear effect of acoustic continuum number on percent "ba" judgments. Both tests were significant (felt "ba": $F(1,99) = 8.54$, $p = .0044$; felt "ga": $F(1,99) = 29.64$, $p < .001$); in each case, felt "ba" judgments decreased as the acoustic syllable shifted from the /b/ to the /g/ end of the continuum. They decreased from 95% "b" judgments to 80% on felt "ba" trials and from 39% to approximately 10% on felt "ga" trials.

Discussion

The most important outcome of this experiment is that a strong cross-modal effect on judged spoken syllables occurred in one cross-modal condition, while at most a marginal effect occurred in the other. In particular, a highly significant effect occurred when mouthed syllables were felt simultaneously with an acoustic presentation of similar syllables, but a marginal or nonsignificant effect occurred when syllables were printed. According to the logic presented in the introduction, we interpret this difference as evidence favoring accounts of speech perception in which perceptual objects are the phonetically significant vocal tract actions that cause structure in light, air and on the skin and joints of the hand. By the same token, we have shown that mere association in experience between sources of information for a syllable at best gives rise to a weak cross-modal influence using this procedure. Shortly we consider an alternative interpretation of these latter findings that Experiment 2 is designed to address. First, however, we consider some other outcomes of Experiment 1.

One interesting outcome is that the cross-modal effect of felt on heard syllables is present in inexperienced subjects in the very first block of trials in which they participate. This suggests that the effect arises in the absence or near absence of experience in which the acoustic and haptic information is paired. Accordingly, we conclude that joint specification of an environmental event does not require specific learning to be effective in perception. We discuss this conclusion further in the General Discussion.

A second interesting finding of the experiment is that the cross-modal effects in the Tadoma condition worked in both directions. Not only did felt syllables affect judgments of the syllable heard, but the acoustic syllable affected judgments of the syllable felt. This suggests considerable integration of the information from the two modalities.

There is an alternative interpretation of our findings in the orthographic condition. Our expectations had been that the Tadoma condition would lead to cross-modal effects, while the orthographic condition would not. Accordingly, we designed the orthographic condition in a way that we thought would optimize its chances of working. That is, we presented the orthographic syllable for a long period of time, guaranteeing that subjects would see the syllable and hence maximizing the number of opportunities for a cross-modal effect to occur. However, the effect of the long-duration presentation may have been different than we expected. Although we asked subjects to look at the screen as they pressed the return key (so that they would see the printed syllable simultaneously with hearing the acoustic syllable), they need not have followed directions, because the syllable remained on the screen until the next trial was initiated. Further, the task of feeling a mouthed syllable is difficult and attention-demanding in a way that the task of looking at a clearly-presented printed syllable is not. One consequence may be that the Tadoma task took attention away from the listening task, leaving the acoustically-signaled syllables less clearly perceived and perhaps more vulnerable to influence of information from another modality. To address these possibilities, we designed a second orthographic condition in which printed syllables were presented briefly and masked.

EXPERIMENT 2

Experiment 2 was designed to assess the effect of attention on the cross-modal influence of printed syllables on acoustic syllables. We hoped to use masking to force subjects to look at the printed syllable at the same time they were listening to the acoustic syllable and to drive performance reporting orthographic syllables down to a level comparable with the nonrandom subjects' ability to report the felt syllables in Experiment 1 (78.6% correct).

Methods

Subjects. Subjects were 12 undergraduate students at Dartmouth College who were native speakers of English with normal hearing and normal or corrected vision.

Stimulus materials. Auditory stimuli were those described in Experiment 1. The visual stimuli were also identical to the visual stimuli in Experiment 1, except they were masked by a row of number signs ("#"). On each trial, a row of four number signs appeared just above the location on the screen where the syllable would be printed. Simultaneous with presentation of the synthetic-speech syllable, either "BA" or "GA" was printed on the screen below the number signs; the printed syllable was covered over after 67 ms by another row of number signs.[3] The mask remained on the screen until the subject hit the key for the next trial.

The 60-trial test orders of continuum members and orthographic syllables used in Experiment 1 were also used here to sequence stimulus presentations.

Procedure. Each subject was tested individually. As in Experiment 1, the endpoints of the continuum were played and identified before the main test to allow subjects to become familiar with the synthetic-speech syllables. They were told that they could pace themselves through the experiment by hitting the return key, and they were instructed not to leave any blanks on the answer sheet, taking a guess if necessary. In addition, as in Experiment 1, subjects were told that acoustic signals and masked visual signals were independently paired on each trial, and, therefore, decisions about them should be made independently.

Results and Discussion

We succeeded in lowering performance on the printed syllables to the level of the nonrandom subjects in the Tadoma condition of Experiment 1. Performance judging orthographic syllables averaged 78.3%, very nearly the same level of performance judging felt syllables found in the Tadoma condition.

The major results of the experiment are depicted in Figure 3. The figure shows no effect of the orthographic syllable on identifications of heard syllables. An analysis of variance on the data in Figure 3 revealed only a significant effect of Continuum number ($F(9,99) = 39.87$, $p < .001$). The effect of orthographic syllable was nonsignificant, with a 4.4% numerical difference between means in the unpredicted direction; the interaction also failed to reach significance.

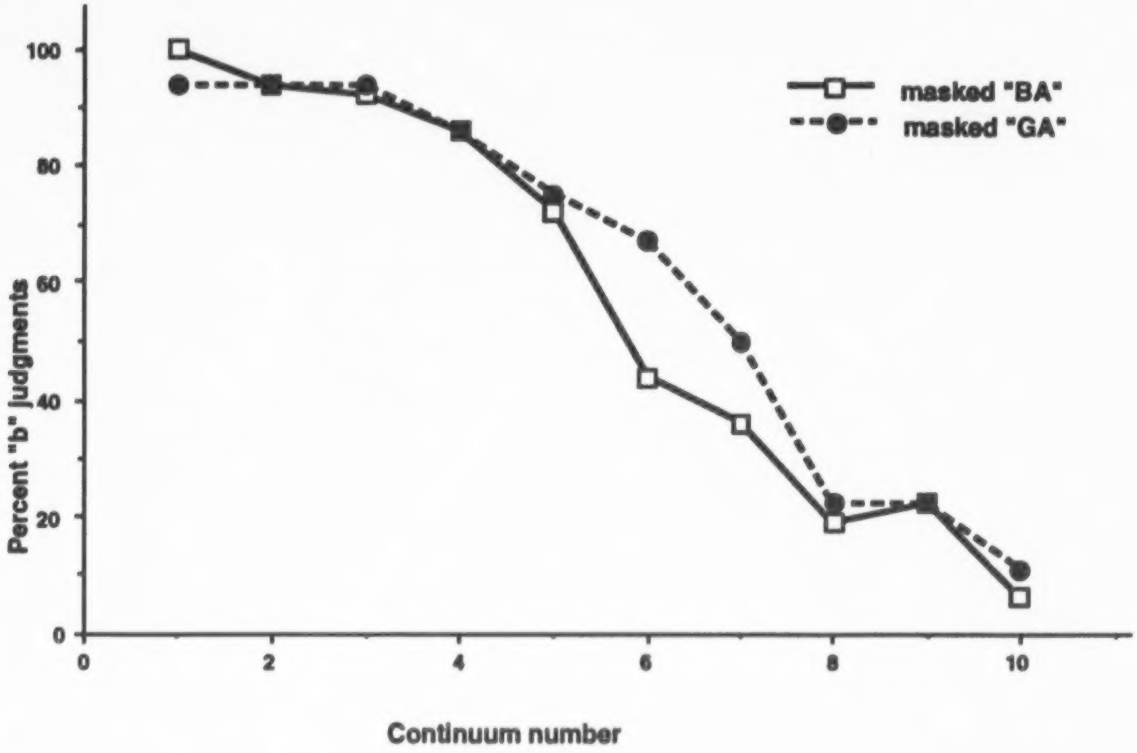


Figure 3. **Percent "b" judgments to acoustic continuum members in the masked orthographic condition of Experiment 2.**

In an analysis of variance including the Tadoma condition of Experiment 1 and the results of the present experiment, the interaction between Condition and Cross-modal syllable was highly significant ($F(1,22) = 17.18$, $p = .0005$).

An analysis of variance on the effect of the acoustic signal on visual judgments yielded two significant effects. The effect of orthographic BA or GA was, of course, highly significant, ($F(1,11 = 319.02$, $p < .001$), accounting for 60.6% of the variance. In addition, there was a marginal effect of acoustic continuum number ($F(9,99) = 1.98$, $p = .049$) that accounted for just 1.4% of the variance. The interaction between continuum and orthographic BA or GA did not reach significance. We performed a trend test on the effect of Continuum number to determine whether decreases in identifications of an orthographic syllable as BA were associated with decreasingly /ba/-like acoustic signals. The analysis, which tested for a linear decrease in "ba" identifications across the continuum, did not approach significance ($F<1$).

We were successful in using masking to decrease identifiability of the orthographic syllables to a level comparable to identifiability of felt syllables among nonrandom subjects in the Tadoma condition. We hoped that this manipulation would increase attention to the visual stimulus while the acoustic signal was being presented and would attract attention away from the acoustically-presented syllables. If these attentional features had been characteristic of the Tadoma condition of Experiment 1, but not of the orthographic condition of that experiment and if those differences had been the operative ones in the different outcomes of those conditions in Experiment 1, then a cross-modal effect should have been strengthened in the present experiment. Instead, the marginal effect in Experiment 1 disappeared completely in Experiment 2; accordingly, we revert to our original interpretation of the difference in outcome. McGurk-like effects occur when information from the two modalities are conjoint, lawful consequences of the same environmental event. They do not occur based only on association in experience.

EXPERIMENT 3

We performed a final experiment, in two phases, to address a variety of methodological concerns raised by reviewers about the haptic condition of Experiment 1. Two concerns were consequences of our having used a loudspeaker rather than headphones to present the acoustic stimuli. Possibly mouthed syllables were whispered, and the influence on heard syllables in that condition was auditory-auditory, not haptic-auditory. In addition, since the model could hear the spoken syllables, perhaps that biased her mouthing actions. While we discount both concerns[4] it was easy to allay them by eliminating the loudspeaker and substituting headphones. Of course, this is likely to reduce the cross-modal influencing effect as do other manipulations that decrease the compellingness of information that the cross-modal signals derive from the same physical event (see, e.g., Easton & Basala, 1982). However, we considered our original effect large enough to survive some reduction due to spatial dislocation of voice and mouth.

A third concern was that we had provided no measure of synchrony between mouthed and audible syllables; we provided a measure, albeit a phenomenal one, in Experiment 3. A fourth concern was with the dropout rate among our subjects. We believed that the poor performance levels on haptic judgments on the part of some subjects in Experiment 1 were due to discomfort with the procedure and the absence of rewards for high performance. Accordingly, in Experiment 3, we instituted rewards.

The most serious concern with the findings, however,was that they are consistent with two interpretations of the cross-modal influence of felt mouthed on heard syllables. One is that the influence reflects a true integration, within a trial, of information from two modalities ostensibly about the same physical event. The other is that information from the two modalities remains unintegrated and that, over trials, subjects sometimes select either the acoustic syllable or the haptic syllable and give it as their response both to the heard and to the felt syllable. We had attempted to discourage such a response strategy by informing our subjects that we had independently and randomly paired mouthed and spoken syllables and that, in consequence, the identification of a syllable in one modality provided no information about the identification of the syllable in the other modality. Moreover, the selection strategy was equally available to subjects in Experiment 2, but it was not adopted there. We interpreted this as evidence that subjects understood, believed, and could follow those instructions. Accordingly, we ascribed the cross-modal influences that occurred in the haptic, but not the masked visual, condition as evidence of integration taking place in one set of conditions and not in the other. However, in Experiment 3, we attempted to provide more direct evidence for integration rather than selection.

Massaro (1987) provides two kinds of information for distinguishing integration and selection in his tests of the effects of visible speaking on identification of spoken syllables. One kind of information is provided by superadditivity of contributions from the modalities to response identifications. In one of Massaro's experiments, on average, subjects correctly identified visible /ba/ presented with no accompanying acoustic signal on 75% of trials. Further, they identified, for example, the most /ba/-like acoustic syllable presented unimodally as "ba" on just over 80% of trials. However, the same acoustic continuum member accompanied by visible /ba/ was identified as "ba" nearly 100% of the time. When that pattern is present to a sufficient degree in individual-subject performances, it rules out one version of a selection strategy, because selection by that strategy will yield an averaging of identification percentages weighted by the relative frequencies with which each modality is selected. The strategy is available to subjects who can process just one modality of information on each trial. In that case, they have just one response to offer, and they may give that response as their identification of syllables presented in both modalities. That strategy will mimic evidence of cross-modal integration of information except that the probability of identifying a given acoustic syllable, say, as "ba" on a bimodal trial cannot exceed a weighted average of the probabilities of identifying each unimodal syllable as "ba." In the example above, performance identifying spoken /ba/ to the first acoustic continuum member could not exceed 80% on bimodal trials, given that the acoustic syllable was identified as "ba" 80% of the time and the visible syllable was identified as "ba" less than 80% of the time on unimodal trials.

We looked for evidence of superadditivity in our data, but we point out here that it does not provide strong evidence against selection.[5] It is implausible to assume that subjects can only process information from one modality at a time. More likely, both modalities provide perceptual information on each trial. A possible selection strategy, then, is, for example, in identifying the heard syllable, to select a response based on the auditory information (so, in the example, identify the syllable as "ba" with probability .8) unless that information is noisy; in that case, use the information from the other modality (in the example, choose "ba" with probability .75). Now the bimodal probability becomes .8 + (1-.8)*.75 = .95, a value higher than either unimodal value.

A second kind of information for integration rather than selection is provided by evidence that the syllables from the modalities blend in some way. For example, given visible /da/ and auditory /ma/, an identification of the spoken syllable as "na" blends place information from the visible syllable and manner information from the acoustic syllable. In Experiment 3, we also looked for evidence of blending.

Methods

Subjects. Subjects were ten undergraduate students at Dartmouth College who participated for course credit. Two subjects of the ten were also paid for high performance identifying haptic and continuum end-point acoustic syllables. All subjects reported normal hearing. Data from three unpaid subjects were eliminated because their performance identifying haptic syllables was at chance.

Stimulus materials. Acoustic stimuli were those used in Experiments 1 and 2. We created four test orders of 96 trials each. Of the 96 trials, 60 were cross-modal audio-haptic syllables as in Experiment 1. In thirty trials, the ten acoustic continuum members were presented by themselves three times each. Six trials were unimodal haptic trials in which the mouthed syllables BA and GA occurred three times each. Unimodal trials are needed to test for superadditivity of bimodal contributions to syllable identification. Trials of the various types were randomized in each test order.

Procedure. Subjects participated in four sessions in each of which one of the test orders was used. The order of the different test sequences was varied across subjects. Generally, instructions were the same as in Experiment 1 except, of course that we told subjects that some trials were not bimodal and that they should just report one syllable on those trials. An additional change in procedure was that subjects now wore headphones over which acoustic syllables were played; the acoustic syllables were, thereby, inaudible to the experimenters. A final change in procedure, meant to improve performance identifying haptic syllables is that we invited subjects to use whatever hand placement on the face made it easiest for them to identify mouthed syllables.

For the first two subjects we ran, we offered a system of payments for high performance on haptic identifications and on identifications of acoustic continuum endpoints. In particular, subjects could receive $0.25 for each percentage point above 75% for haptic identifications and another $0.25 for each percentage point above 75% for identification of acoustic endpoints. Subjects could, therefore, earn a maximum of $50.00 in all across the four sessions.

We also asked these same subjects to give us a final judgment on each trial in addition to identification of heard and/or felt syllables. We asked them to tell us whether the syllables in the two modalities were synchronous or not. In particular, we told them that they should report "E" if the mouthed syllable led the acoustic syllable, an "S" if it was synchronous and an "L" if the mouthed syllable lagged the acoustic syllable.

Both of these latter procedures were abandoned after two subjects were run. The system of payments was abandoned because performance on haptic unimodal syllables was near perfect, making it impossible to test for integration by looking for superadditivity. (One subject earned $43.00 of the maximum possible of $50.00; the other earned $39.50. On unimodal haptic trials, performance averaged 92% correct—too high to look for superadditivity of crossmodal influences on syllable identification.) The assessment of simultaneity was abandoned because, although we did not bias subjects to expect simultaneity, they reported simultaneity virtually all of the time (on 97% of trials for one subject and 98% for the other).

Because we feared that superadditivity would remain difficult to test for and since we have learned that it does not provide a strong test of integration anyway, we made a final change in procedure that we hoped would enable us to test for integration by testing for blend responses. For the remaining subjects, we opened the response inventory in the following way. As we had done in Experiments 1 and 2 and in running the first two subjects of Experiment 3, we played the acoustic-continuum endpoints to subjects before they began the experimental test. In addition, for the remaining subjects, we told them that there were ten syllables in all and that the other eight ranged between the /ba/ and /ga/ syllables they had just heard. We played them continuum member 5 as an example of an ambiguous syllable. We told subjects that we did not know how listeners would identify those intermediate sounds; they might sound like ambiguous /ba/s and /ga/s or they might sound like other consonants to them, possibly /ða/, /va/ or /da/. We asked them always to report the sound they heard, whether it was /ba/, /ga/ or some other consonant sound. We also gave them to understand that, haptically, while they should feel some clear instances of mouthed /ba/s and /ga/s, sometimes they might feel other consonant-initial syllables. On judgments of felt syllables, they should always report the syllable they experienced, whether it was /ba/, /ga/ or some other consonant-initial syllable.

Because, in fact, our procedural changes made little difference in subjects' performances, for most analyses, we have pooled the findings on all subjects who completed the four sessions. (Three subjects of the eight unpaid subjects were dropped after one or two sessions for chance performance on haptic trials. Thus, after abandoning our system of payments, performance that had been at ceiling on unimodal haptic trials and was very high on bimodal trials, reverted to a level just 3% higher than its level in Experiment 1. We believe that subjects can do the haptic task if they are motivated to attend.)

Results

Figure 4 presents the judgments of heard syllables on the auditory only (top) and cross-modal (bottom) trials of Experiment 3. In an analysis of variance on cross-modal trials, with factors Continuum number and Haptic syllable, both main effects were significant (continuum: $F(9,54) = 74.60, p < .0001$; haptic syllable: $F(1,6)= 16.29, p = .007$). The interaction did not approach significance ($F<1$). While the effect of haptic syllable was numerically smaller than in Experiment 1, perhaps due to the spatial dislocation of cross-modal syllables, it is present on nearly every syllable along the continuum and is highly significant.

Another reason for the reduced cross-modal effect as compared to its magnitude in Experiment 1 might be that subjects learned over sessions to divide their attention across sessions. To look for effects of practice, a further test was performed, now on the percent "ba" responses averaged across the continuum and now including auditory-alone trials, with session as an independent variable. In the analysis, there was an effect of haptic syllable (BA, GA, none; $F(2,12) = 4.81, p = .029$) and an effect of block ($F(3,18) = 3.20, p = .048$); however, the interaction did not approach significance ($F<1$). As in Experiment 1, "ba" responses to auditory-alone syllables (49.5%) fell between responses to syllables accompanied by mouthed BA (54%) and responses to syllables accompanied by mouthed GA (49%). Now, however, they fell closest to (and very close to) performance on mouthed GA trials, rather than mouthed BA trials as they had in Experiment 1. The effect of sessions occurred because the percentage of "ba" responses declined monotonically from session 1 (54.8%) to session 4 (48.7%). However, the size of the cross-modal effect did not change across sessions, and the numerical change was in the direction of an increased, not a decreased, effect with practice.

We looked also at the effect of acoustic syllables on haptic judgments. We compared the percentage of "ba" judgments to mouthed BA and GA when they were accompanied by acoustic syllables from the /ba/-end (items 1-5) or the /ga/-end (6-10) of the continuum. Table 1 gives the results. For both mouthed syllables, identification of the felt syllable as "ba" was more likely when the mouthed syllable was produced synchronously with a /ba/-like acoustic syllable than when it was produced with a more /ga/-like syllable. An analysis of variance was performed on the data in Table 1, arc-sine transformed to eliminate variance differences due to approaches to ceiling for mouthed BA at the /ba/-end of the continuum and to floor of mouthed GA produced at the /ga/-end of the continuum. The effect of haptic syllable was, of course, highly significant ($F(1,6) = 45.69, p = .0007$). The effect of acoustic syllable was also significant ($F(1,6) = 6.42, p = .04$) as was the interaction ($F(1,6) = 9.72, p = .02$). The interaction occurred because the effect of continuum was 28.7% for haptic GA but only 7.7% for haptic BA, where performance was very close to ceiling.

We turn now to tests of integration. Generally, the test for superadditivity was not applicable to subjects, because their performance identifying haptic syllables was at ceiling on unimodal trials. In particular, none of our seven subjects was less than 92% correct in identifying haptic BA as such on haptic alone trials; five subjects made no errors. Obviously, we cannot look for superadditivity of effects of haptic and acoustic information using haptic- and acoustic-alone trials to predict performance on bimodal trials. We do not know why Massaro's subjects performed relatively poorly (75% accurate in his Figure 7, p. 65 of Massaro, 1987) identifying /ba/ in a video-only condition, so that evidence of superadditivity could be sought. By the same token, three of our seven subjects made no more than 8% errors identifying /ga/ on haptic alone trials. We looked at performances of the remaining four subjects.

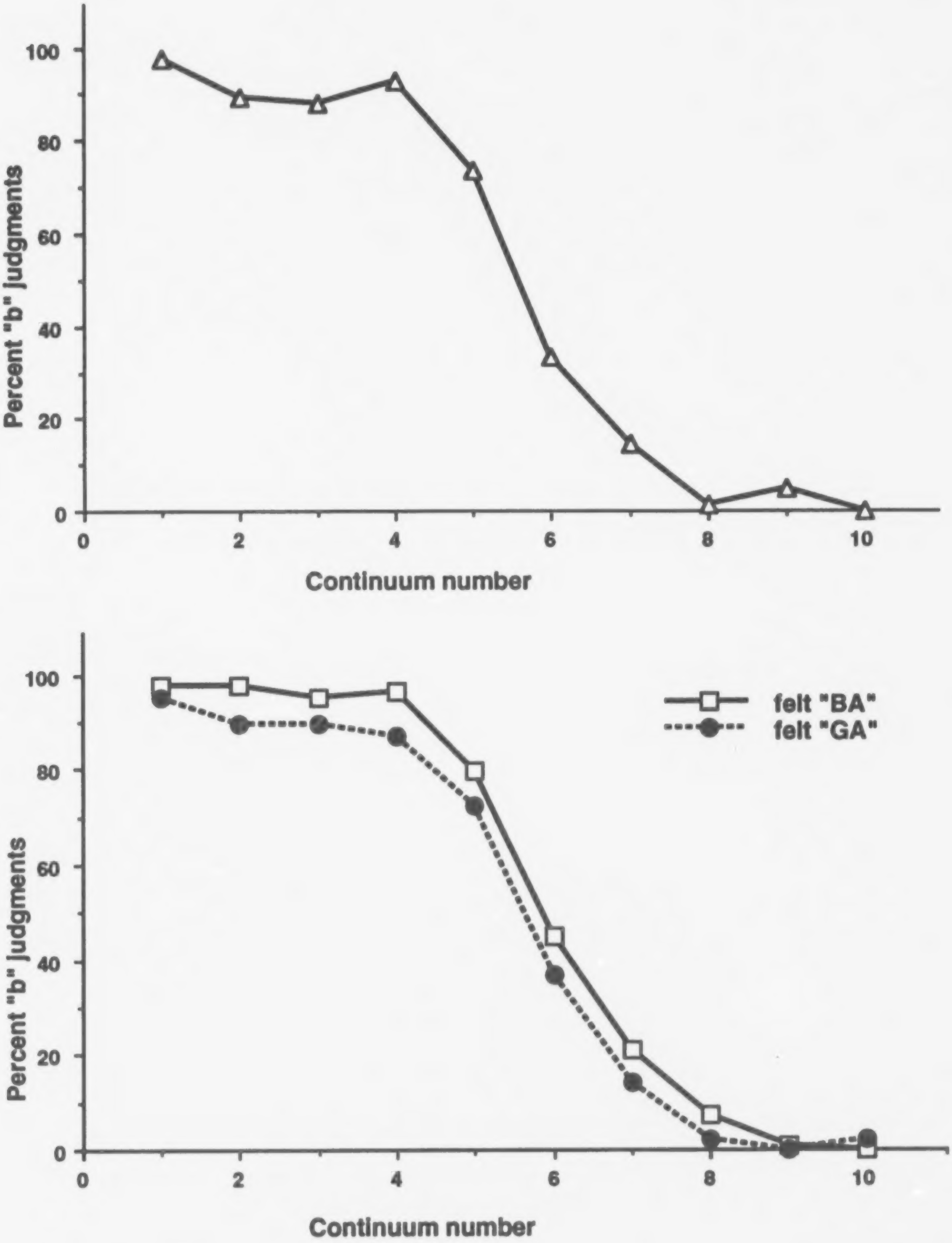


Figure 4. **Percent "b" judgments to acoustic continuum members on the auditory-alone trials (top) and cross-modal trials of Experiment 3.**

Table 1. *Percent "ba" responses to haptic BA and GA on trials in which mouthed syllables were produced simultaneously with synthetic syllables from the /ba/-end (continuum numbers 1-5) and /ga/-end (6-10) of the continuum.*

	/ba/-end	/ga/-end
Haptic BA	95.4	87.6
Haptic GA	35.7	7.0

Of them, two showed some evidence of superadditivity. Across haptic and acoustic identifications on trials involving GA as a mouthed syllable, there are 20 opportunities to show superadditivity of haptic-alone and acoustic-alone contributions to bimodal identification of the mouthed or acoustic syllables. Two of the four subjects showed superadditivity on six of the 20 trials; of the remaining two subjects, one showed superadditivity on two trials and one on none. We do not know how to evaluate this outcome statistically. (In Massaro's grouped data (p. 65) four of nine continuum members show superadditivity relative to video-alone /ba/, while none do so relative to video-alone /da/ where performance is at ceiling; in his sample individual subject data, two continuum members show superadditivity relative to video-alone /ba/ while none do relative to video-alone /da/, where, again, performance is at ceiling. Massaro evaluated superadditivity by testing his model, in which bimodal influences are integrative and hence superadditive against a competing, selection, model. Since we reject Massaro's model as a source of cross-modal haptic influences on auditory speech perception and vice versa, we did not copy his procedure.)

A second test for integration involves tests for blends of information presented in the different perceptual modalities. To enable us to test for blends we opened the response inventory for subjects after the first two subjects that we tested in Experiment 3. However, just one subject of the remaining five used a response other than "b" or "g" on more than a dozen occasions. The one subject used "va" as a frequent response both for heard- and for felt-syllable judgments. Here we consider the patterning of his "va" responses.

Since /ba/ and /ga/ are articulatory extremes in English, almost any other consonantal response other than "b" or "g" is intermediate between /ba/ and /ga/ and hence looks, on the surface, like a blend. To attempt to determine whether the subject who gave frequent "va" responses was giving true blends or, alternatively, was simply guessing giving another consonant name, we looked for a systematic patterning in "va" responses. We reasoned that, if "va" responses are blends, then when BA is the mouthed syllable, "va"s should be given as an identification of either the heard or the felt syllable more frequently when acoustic stimuli are at the /ga/-end of the continuum than when they are at the /ba/-end. That is, as the haptic syllable pulls the response /ba/-ward, the acoustic syllable pulls it /ga/-ward yielding a blend response. Similarly, when GA is the mouthed syllable, "va" identifications should be more frequent when acoustic syllables are at the /ba/- than at the /ga/-end of the continuum. For purposes of the analysis, we eliminated middle-continuum syllables 5 and 6 from the analysis and examined identifications of acoustic and haptic syllables when acoustic syllables were /ba/-like (continuum numbers 1-4) or /ga/-like (7-10). Table 2 shows the results for the single subject who gave a considerable number of "va" responses.

Table 2. *One subject's frequency of "va" responses to acoustic (left) and haptic (right) syllables on bimodal trials, when the synthetic syllables were from the ba- or the ga-ends of the continuum. Middle continuum numbers (5 and 6) were disregarded.*

	Judged heard syllable		Judged felt syllable	
	ba-end	ga-end	ba-end	ga-end
Haptic BA	5	6	3	5
Haptic GA	10	1	14	4

His judgments of both acoustic and haptic syllables show the predicted pattern. That is, when BA is the mouthed syllable, "va" identifications of both the acoustic and the haptic syllables are more likely when the acoustic syllable is from the /ga/- than from the /ba/-end of the continuum. The pattern is reversed when the mouthed syllable is GA. Neither table has sufficiently high frequencies for a Chi Square analysis; however, a pooled table does. In an analysis, then, of the likelihood of a "va" identification of either a haptic or acoustic syllable when haptic and acoustic syllables pull in the same or opposite directions, the Chi Square value is highly significant ($\chi^2 = 8.54$, $p = .0036$). Based on this analysis, we can conclude that there exists at least one person for whom the influence of haptic on acoustic judgments and vice versa is a true integration and not a selection. Possibly, two other of the seven subjects show evidence for integration in the form of superadditivity of cross-modal influences as well.

GENERAL DISCUSSION

Our findings in the Tadoma condition of Experiment 1 and in Experiment 3 suggest that perception need not be based on covert anticipations of the range of sensory cues that may be experienced from stimulation nor on associations between those cues and representations of their environmental causes. Together with results of the orthographic conditions of our experiments, the findings suggest, indeed, that stored associations are not sufficient for cross-modal integration of information for an event and that association in the world, specifically due to (ostensible) joint causation by a common environmental event, is required.

There is another indication, in our view, that perception of speech syllables does not require prior existence in memory of a prototype or some other way of associating sensory cues to representations of perceivable events. In a McGurk procedure, when the acoustic syllable is /da/ while the model mouths /ba/, listeners frequently report having heard /bda/ (e.g., Massaro, 1987). In Massaro's fuzzy-logical model of perception (FLMP), a syllable is reported by listeners if it is represented by a prototype in memory, and if evidence consistent with that prototype is stronger than evidence consistent with other prototypes as determined by Luce's choice rule (1959). In discussing the definition of prototypes in the model, Massaro refers to a prototype for /bda/ (1987, p. 128). But how could a hearer acquire a prototype for /bda/? /bda/ is not a possible syllable in English. Indeed, it violates the language-universal sonority constraint (roughly that consonants in a within-syllable cluster must increase in vowel-likeness toward the syllable's vowel; e.g., Selkirk, 1982). Listeners will not have experienced /bda/ prior to the experiment; nor will evolution have provided a prototype for a universally disallowed syllable.[6] The only possibility, it seems, is that prototypes can be constructed "on the fly" in the experiment. That is, there must be a way for the perceiver to decide that no existing prototype is sufficiently supported by the evidence in stimulation. In that case, a new prototype is established and named. However, if the information in stimulation is sufficient to reveal that a new prototype is needed and to name the prototype /bda/, then, it seems, the prototype itself is not needed for perception. Rather, it *depends* on perception for its establishment.

We also doubt that our outcome is compatible with the motor theory. Liberman and Mattingly (1985) invoke an innate vocal-tract synthesizer that subserves both speech production and, using something like analysis-by-synthesis, speech perception as well. There is reason to suppose that an innate synthesizer would have anticipated the possibility of receiving optical as well as acoustic information about vocal-tract gestures since both of those information sources are commonly available in the environment of listeners; ability to exploit the information sources might have adaptive significance. However, there is no reason to suppose that selection would have favored evolution of a synthesizer that anticipated receiving haptic information provided by the hands of a listener. More generally, we doubt that any explanation for the Tadoma effect can work that depends on the significance of the haptic information being appreciated either because the significance has been learned or because it is known innately.

How can a new environmental occurrence be perceived, or likewise in our experimental situation, how can a familiar occurrence be signaled effectively, in part by novel proximal stimuli? We follow Shaw, Turvey, and Mace (1982) in concluding that perceptual experience is fundamentally knowledge acquired due to the "force of existence" of events in the world. In our terms (and not necessarily those of Shaw et al), environmental events causally structure media such as light, air and the skin and joints of a perceiver. To the extent that the structure is distinctive to its causal source, it can serve as information about the source. The information can inform without prior familiarity with it because of the causal chain that supports perception. Stimulation caused by an environmental event has causal effects on sensory receptors so that its structure is, in part, transmitted to a perceptual system. By hypothesis, the perceiver comes to know an event in the environment via its impact on the perceptual systems as transmitted by proximal stimulation.

In that sense, perception may not be a particularly intellectual endeavor at all; it may be more analogous to motor accommodations to, or motor exploitations of, physical forces exerted on the bodies of actors. In the analogy, proximal stimuli at the sense organs are the informational

analogs of the physical forces impinging on the body of an actor. We need not learn what haptic consequences of environmental events mean in order to perceive their source any more than we have to learn what gravity means in order to be affected by it in a coherent way. The role of learning, then, may be to change the perceiver/actor's state of preparedness for receiving and exploiting the "forces," both physical and informational, that the world has to offer, not to discover what environmental events the forces signal. That is given in the stimulation. Even in the absence of relevant learning by the perceiver/actor, the environment exerts its same forces that causally affect the body, including the perceptual systems.

REFERENCES

Campbell, R. (1989). Seeing speech is special. *Behavioral and Brain Sciences, 12,* 758-759.

Chomsky, C. (1986). Analytic study of the Tadoma method: Language abilities of three deaf-blind subjects. *Journal of Speech and Hearing Research, 29,* 332-347.

Easton, R., & Basala, M. (1982). Perceptual dominance during lipreading. *Perception and Psychophysics, 32,* 562-570.

Erber, N. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12,* 423-425.

Ewertsen, H., & Nielsen, H. B. (1971). A comparative analysis of the audiovisual, auditive and visual perception of speech. *Acta Otolaryngolica, 72,* 201-205.

Fowler, C. A. (1986) An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3-28.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton-Mifflin.

Keller, H. (1903). *The story of my life.* New York: Doubleday.

Liberman, A. (1982). On finding that speech is special. *American Psychologist, 37,* 148-167.

Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461.

Liberman, A., & Mattingly, I. (1985). The motor theory revised. *Cognition, 21,* 1-36.

Luce, D. (1959). *Individual choice behavior.* New York: Wiley.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception and Psychophysics, 24,* 253-257.

Massaro, D. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. (1989). Multiple book review of *Speech perception by ear and eye: A paradigm for psychological inquiry. Behavioral and Brain Sciences, 12,* 741-755.

Massaro, D., Cohen, M., & Thompson, L. (1988). Visible language in speech perception: Lipreading and reading. *Visible Language, 22,* 8-31.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Rosenblum, L. (1987). Towards an ecological alternative to the motor theory of speech perception. *Perceiving-Acting Workshop Review, 2,* 25-28.

Schultz, M., Norton, S., Conway-Fithian, S., & Reed, C. (1984). A survey of the use of the Tadoma method in the United States and Canada. *Volta Review, 86,* 282-292.

Seidenberg, M., & Tanenhaus, M. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 546-554.

Selkirk, E. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations, Vol 2* (pp. 337-384). Dordrecht, The Netherlands: Foris Publications.

Shaw, R., Turvey, M. T., & Mace, W. (1982). Ecological psychology: The consequence of a commitment to realism. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes, 2* (pp. 159-226). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212-215.

Summerfield, A. Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3-51). London: Lawrence Erlbaum Associates.

Summerfield, A. Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology, 36A,* 51-74.

Tanenhaus, M., Flanigan, H., & Seidenberg, M. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory and Cognition, 8,* 513-520.

FOOTNOTES

[*] *Journal of Experimental Psychology: Human Perception and Performance, 17,* 816-828 (1991).

[†] Also Dartmouth College.

[††] Dartmouth College.

[1] As we discuss later, we attempted to bias the experiment in favor of the orthographic condition by presenting the printed syllable for a long period of time. Possibly, however, this was a design flaw in that the printed syllables failed to demand as much attention as the haptically-perceived syllables. We provided a more attention-demanding orthographic condition in Experiment 2.

[2] In an analysis including all 22 subjects, the main effects of Continuum number and of Cross-modal syllable remain highly significant, while the critical interaction is marginally significant ($F(1,21) = 4.16$, $p = .052$). The difference in heard "b" responses as a function of feeling mouthed "ba" versus "ga" is 12.6%; as a function of seeing printed "ba" versus "ga" it is 5.6%. We performed this test including random subjects for two reasons. One was a discomfort with eliminating nearly half our sample of subjects in the main analyses. A second was a sense that, because subjects made the "felt" judgments second, some forgetting might have occurred. If so, their performance on the felt judgments may have underestimated both their true ability to identify felt "ba" and "ga" and the influence that the felt syllable might have had on the heard syllable.

[3] The stimulus onset asynchrony (SOA) was determined by pilot testing subjects until we found an SOA at which performance identifying the spelled syllables was approximately that of the haptically-identified syllables of Experiment 1. The SOA at which performance approached that on the Tadoma condition of Experiment 1 was 67 ms. Occasional trials may have had SOAs of 83 ms, however, because we were unable to control target and mask presentations relative to the occurrence of screen refreshes.

[4] In Experiment 1, only supraglottal actions of the model mimicked those of naturally-spoken /ba/ and /ga/; there was no

BEST COPY AVAILABLE

channeling of air through the oral cavity and, accordingly, there was no whispering. As for biasing effects on the model of hearing the synthetic syllables, two factors precluded that. First, to make the movements, particularly of /ba/, synchronous with the synthetic syllable—a condition we presumed important to integration (cf. Easton & Basala, 1982)—closing movements for the consonant had to be initiated *before* the onset of the signal so that articulatory release, where the onset of acoustic energy begins for stop consonants, co-occurred with the stop burst and vowel onset. Second, however, were the model to delay mouthing the syllable until the identity of the spoken syllable was detectable, no other behavior than following instructions to mouth the syllable printed on the CRT screen could have led subjects to shift their judgments of the heard syllable in the direction of the mouthed syllable.

[5] We thank George Wolford for pointing this out.

[6] While in rapid speech production, certain "be-" words (such as "beneath" or "become") may by reduced so that the first vowel is inaudible, there are no entries in the dictionary whereby such reduction would lead to a /bda/ syllable. (Readers of the manuscript have suggested the words "bidet," "bedazzle," "bedecked," and "bedevil," however, they should recall that Massaro's memory representations are *syllable* prototypes, not consonant prototypes. Were the schwa vowels dropped in the foregoing words, the remaining syllables would be /bdey/, /bdaez/, /bdɛk/ and /bdɛv/ not /bda/ (/a/ being the first vowel in "father").

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 81-86

Recovering Tube Kinematics Using Time-varying Acoustic Information

R. S. McGowan

Formant frequency trajectories are used to optimally fit the kinematics of a modified twin-tube. An entire articulatory trajectory is fit in a single optimization, because an articulatory trajectory is modeled as a parameterized function of time.

1. INTRODUCTION

The inverse mapping between acoustics and articulation has received considerable attention in the last twenty-five years. The focus has been on mapping static spectral variables onto static vocal tract shapes, with resulting ambiguity in the mapping. Ambiguities were noted in the work of Atal, Chang, Mathews, & Tukey, 1978, where the articulatory positions of the vocal tract model were varied to fit formant frequency data; in the work of Flanagan, Ishizaka, & Shipley, 1980, using an optimization procedure based on spectral information and cepstral matching to find vocal tract area functions, as well as subglottal pressure and laryngeal parameters; and the work of Levinson & Schmidt, 1983, using a gradient search optimization to relate articulator positions to LPC envelopes.

Two ways of overcoming inverse mapping ambiguities suggest themselves: either decrease the number of articulatory degrees of freedom, or increase the amount of acoustic data. One procedure to decrease the number of articulatory degrees of freedom took account of the continuity of vocal tract tube shapes in short time intervals (Kuc, Tuteur, & Vaisnys, 1985; Mermelstein, 1967; Shirai & Kobayashi, 1986). This seemed to help relieve ambiguity, but the optimizations were performed at each time sample, making the inclusion of the continuity constraint inefficient. In the method examined here, the kinematics of the articulators were parameterized as functions of time, and the optimization was performed over time spans corresponding to a single parameterization, thus the continuity constraints were automatically incorporated. Because the time spans were longer than a single time sample, there was a span of acoustic data that was used in the optimization, thus the number of degrees of freedom in the data was also increased.

2. Method

The acoustic data consisted of up to three formant frequency trajectories that were generated using a modified twin-tube model (Fant, 1960). In the modification considered here, a third tube, a constriction tube, was placed between front and rear tubes of the twin-tube model (Figure 1). There were five articulatory variables: front tube area, constriction tube area, rear tube area, rear tube length, and constriction tube length. The front tube length was determined by the restriction that the total tube length be 17 cm. The con-

---

This work was supported by NIH Grant HD-01994 to Haskins Laboratories.

striction tube could change area through time, thus opening and closing the tube between the front and rear tubes. The constriction area was parameterized as an exponential function of time. The maximum area of the constriction was assumed to be the average of the front and rear tube areas, and the minimum was zero, corresponding to complete constriction. As a result there were five articulatory kinematic parameters: the four constant articulatory variables, and the exponential growth factor for the change in constriction area (Figure 2).

The modified twin-tube model was used for both the synthesis of formant frequency data and as a model vocal tract for articulatory kinematic parameter recovery. The relationship between the acoustic variables and the articulatory variables was given by the model function. This function was written as an implicit relation between the formant (resonance) frequencies and the articulatory variables. Thus, if the constriction area was given a trajectory, either opening or closing, it is possible to compute the corresponding formant trajectories using numerical root-solving techniques.

Preliminary work has been done on recovering articulatory kinematic parameters from synthesized formant frequency trajectories using the modified twin-tube model using a least-squares criterion. The iterative least squares was performed using the simplex method (Press, Flannery, Teukolsky, & Vetterling, 1986). The simplex method was a conservative choice because it did not require numerical computation of a generalized inverse, as, say, the Levenberg-Marquardt algorithm did, thus reducing the possibility of numerical instability in this initial study. However, the simplex method was very slow and could be replaced with more sophisticated optimization algorithms. When the experimenter executed the program written for inverse mapping he was asked to specify the constriction length and was given the option of specifying either the front or rear tube areas. If neither of these was specified, then the optimization was performed to find four parameters: the front and rear tube areas, the rear tube length, and the exponential time constant. If one of the areas was specified, then the optimization was performed on three parameters, and if both areas were specified, then two parameters entered into the optimization: rear tube length and the exponential time constant. Because the optimization procedure was an iterative procedure that could be trapped in local minima, the simplex method was run from several initial starting places in the articulatory kinematic parameter space. The search from any of these initial starting places would terminate if the cost function was less than a given tolerance, if there was little relative change in the value of the cost function from one step to another, or if a maximum number iterations was attained.

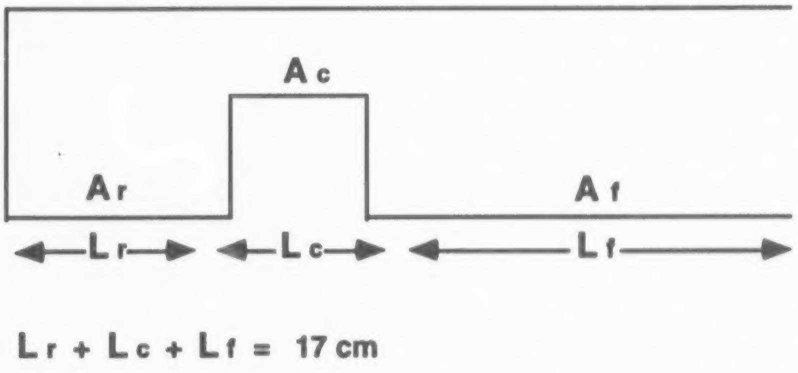


Figure 1. Modified twin tube.

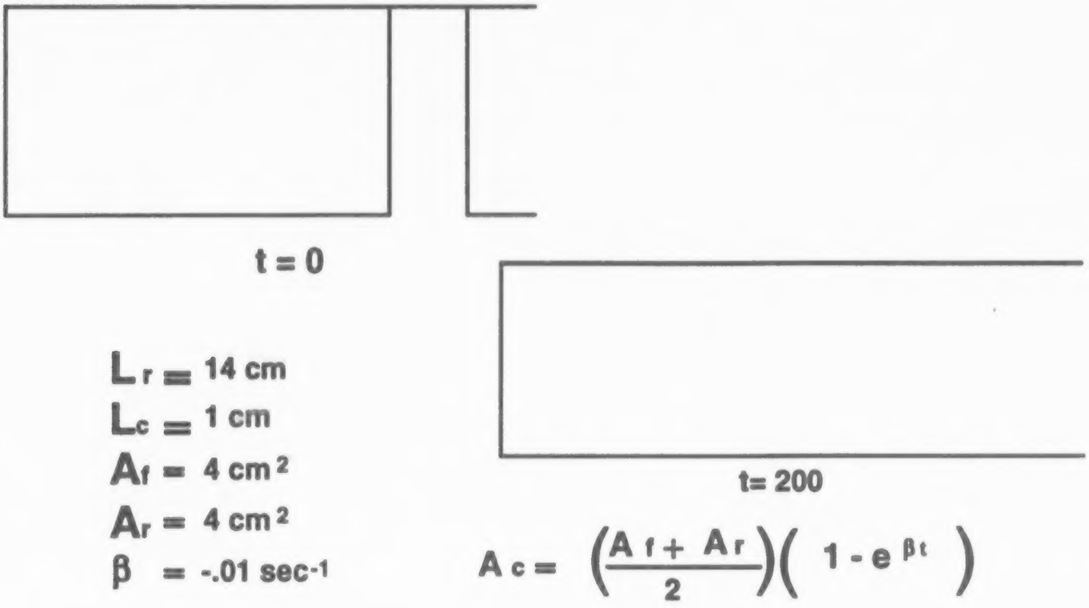
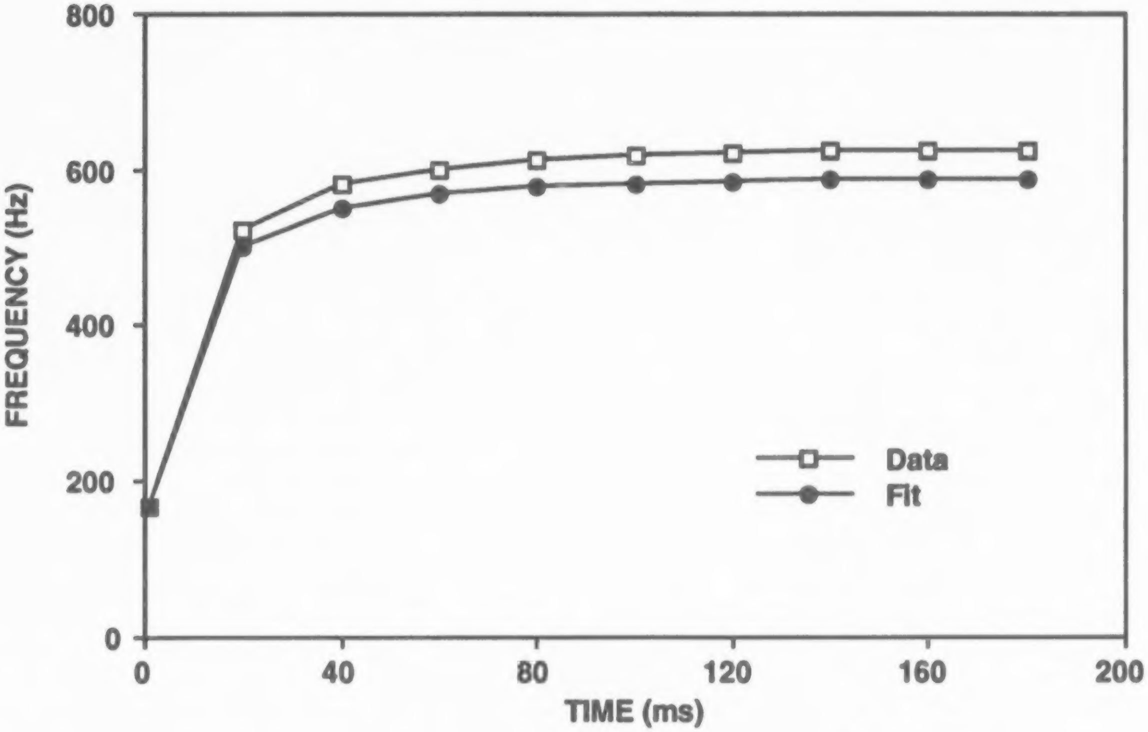

t = 0
t= 200
L_r = 14 cm
L_c = 1 cm
A_f = 4 cm^2
A_r = 4 cm^2
β = -.01 sec^{-1}
$A_c = \left(\frac{A_f + A_r}{2}\right)\left(1 - e^{\beta t}\right)$

Figure 2. **Articulatory kinematic parameters.**

The ideal cost function was the sum of squares of the differences over time in each formant frequency between those given by the data and the values that would be produced by the modified twin-tube model given the articulatory kinematic parameters. To have found the value of this cost function at every iteration, many formant frequencies, corresponding to a given set of articulatory kinematic trajectories, would have had to have been found. This would have involved applying root-solving techniques to the model function many times (40 times for each formant at a rate of 200 Hz for 200 ms). Accordingly, the sum of the squares of the model function evaluated at each data formant frequency was used as an alternative cost function. This appeared reasonable because it is a necessary condition that this function, being an implicit relation between formant frequency and articulatory variables, be identically zero, if the original cost function is zero.

3. Results and Conclusion

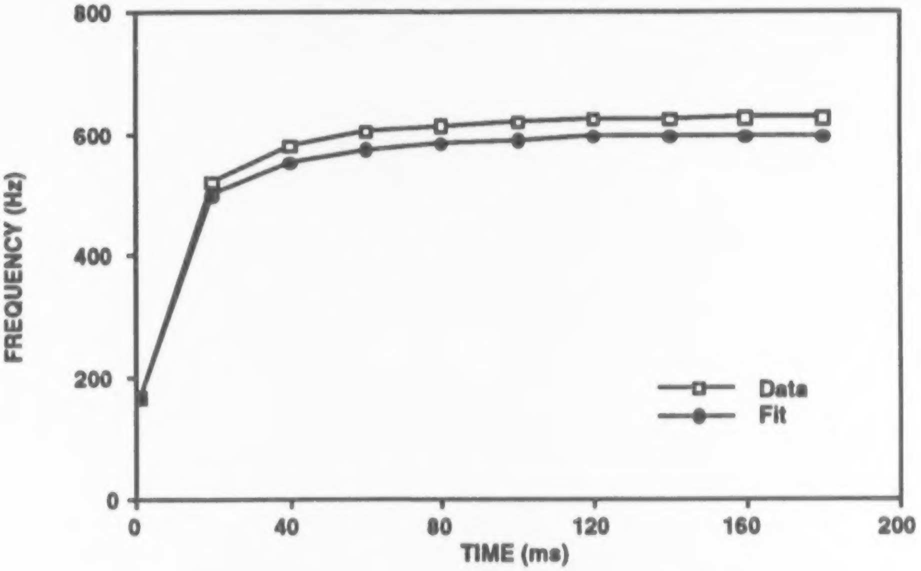
In the modified twin-tube model, the feasibility of fitting rear tube length and exponential time constant was tested using the first formant frequency trajectory only, as well as with three formant trajectories. The feasibility of fitting four parameters, the rear tube area, front tube area, rear tube length, and exponential time constant using one and three formant frequency trajectories was also tested. As one would expect, the method did better in fitting two parameters than it did in fitting four parameters. A counter-intuitive result is that the method seemed to have worked better with one formant (e.g., Figure 3) than it did with three (not shown), or with less information than more. (The program was completely unsuccessful at fitting four parameters given three formant frequencies.)



	rear area (cm^2)	front area (cm^2)	constriction length (cm)	rear tube length (cm)	growth factor (sec^{-1})
Data	2.	4.	2.	8.	-.01
Fit	1.684	4.178	fixed	9.7501	-.0111

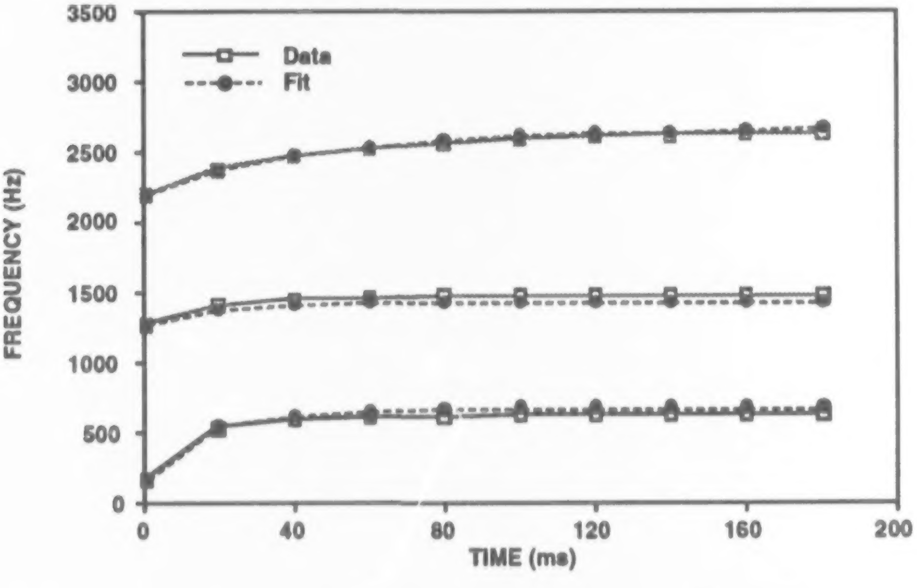
Figure 3. One resonance frequency trajectory, implicit function minimization.

It was felt that something of the original cost function involving the squares of the differences between formant frequency data and those which would be produced with a given set of articulatory kinematic parameters had to be preserved to get better results. Instead of root-solving for all the formant frequency values corresponding to a given set of articulatory kinematic parameters, root-solving was performed only at the beginning, middle, and end of a trajectory for each iteration of the least-squares procedure. (For example, there were nine root solves for three formants.) The sum of squares of the differences between these frequency values and their corresponding data points were added to the sum of squares of the model function evaluated at all the data points to form a hybrid cost function. This seemed to have alleviated the counter-intuitive result of doing more poorly with three formants (Figure 5) than with one (Figure 4). Also, it was possible to fit the four parameters using three formant trajectories (Figure 5).



	rear area (cm2)	front area (cm2)	constriction length (cm)	rear tube length (cm)	growth factor (sec^{-1})
Data	2.	4.	2.	8.	-.01
Fit	2.587	4.181	fixed	10.659	-.0095

Figure 4. One resonance frequency trajectory, implicit function and frequency difference minimization.



	rear area (cm2)	front area (cm2)	constriction length (cm)	rear tube length (cm)	growth factor (sec^{-1})
Data	2.	4.	2.	8.	-.01
Fit	1.284	3.748	fixed	8.0175	-.0081

Figure 5. Three resonance frequency trajectories, implicit function and frequency difference minimization.

The problem with using just the sums of squares of the model function in the cost function was that local minima appeared that were not close to the articulatory kinematic parameters that produced the data. By adding some explicit information to the cost function these superfluous minima no longer hindered the algorithm.

REFERENCES

Atal, B. S., Chang, J. J., Mathews, M. V., & Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique. *Journal of the Acoustical Society of America, 63,* 1535-1555.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Flanagan, J. L., Ishizaka, K., & Shipley, K. L. (1980). Signal models for low bit-rate coding of speech. *Journal of the Acoustical Society of America, 68,* 780-791.

Kuc, R., Tuteur F., & Vaisnys, J. R. (1985). Determining vocal tract shape by applying dynamic constraints. *Proceedings of ICASSP,* 1101-1104.

Levinson, S. E., & Schmidt, C. E. (1983). Adaptive computation of articulatory parameters from the speech signal. *Journal of the Acoustical Society of America, 74,* 1145-1154.

Mermelstein, P. (1967). Determination of the vocal-tract shape from measured formant frequencies. *Journal of the Acoustical Society of America, 41,* 1238-1294.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). *Numerical recipes*. Cambridge: Cambridge University Press.

Shirai, K., & Kobayashi, T. (1986). Estimating articulatory motion from speech waves. *Speech Communication, 5,* 159-170.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 87-92

Motor Levels of Speech Timing: Evidence from Studies of Ataxia*

Fredericka Bell-Berti[†] and Claude Chevrie-Muller[††]

INTRODUCTION

"Fluency" is essentially a perceptual descriptor of the rhythmic patterns of speech, and "dysfluency" is generally taken to be a descriptor of a particular kind of disruption of these rhythmic patterns. We think that it may be profitable to expand this view, and employ a more general definition of "dysfluency"—one that includes all disruptions of the rhythmic patterns of speech. Exploring these various types of "dysfluency" should provide insights into the origins of the different components of speech rhythmic patterns. That is, such study ought to enlighten our understanding of the role of neural, and other physiological, mechanisms in generating these rhythmic (i.e., timing) patterns and disrupting them.

We have been looking at speech timing patterns in the speech of dysarthric individuals who suffer from Friedreich's Ataxia. Darley, Aronson, and Brown's (1975) perceptual descriptions of the rhythmic aspects of speech timing in ataxic dysarthria include excess and equal stress, prolonged intervals, and slow speaking rate. Our particular interest is in quantifying the description "slow speech," which might be speech to which pauses have been added (between syllables, or even between segments). Alternatively, it might be speech in which the durations of segments are increased, either uniformly or differentially across segment type (vowel or consonant) or across an utterance. Our purpose was to identify the nature of the durational relations in the speech of dysarthric subjects, especially utterance-level timing patterns.

As we know, many factors other than speaking rate influence the durations of speech segments in the speech of persons without neurological or other disorder (see Klatt, 1976, for an extensive review). So, for example, the number of syllables in an utterance influences segment duration: "compensatory shortening" refers to the fact that segments occurring early in an utterance are progressively shortened as the length of the utterance is increased by adding syllables (e.g., O'Shaugnessy, 1974). In addition, the position of segments within an utterance influences their duration: "terminal lengthening" refers to the increased duration of segments when they occur in syllables in utterance-final position, in contrast with their durations when they occur early in an utterance (e.g., Lindblom & Rapp, 1973). What is more, these durational patterns may be perceptually important to listeners: Klatt and Cooper (1975) have shown that listeners expect a longer duration for words in phrase-final position both for sentence-internal and sentence-final phrase boundaries, and Nakatani and Schaffer (1978) have found that relative syllable durations are important cues to the location of syllable stress and word boundaries in English.

Simply identifying situations in which segment durations change, though, does not reveal the linguistic or physiological levels at which those durational changes are controlled; i.e., such description does not tell us directly about the relations among phonological, semantic, syntactic, neurological, muscular, and mechanical components of the speech system. However, there are a number of hypotheses about the temporal control system of speech, hypotheses that generate testable predictions about speech timing. For example, Bonnot's model of speech-timing-control levels (1989) postulates two levels of control. First, there is a motor planning level (that results in what Tatham (1970) has referred to as "notional time"), whose

---

This work was supported by the French Ministère de Recherche et Technologie, St. John's University, and NIH Grants DD-00121 and RR-05596 to Haskins Laboratories.

output temporal pattern is the result of linguistic and motor programming interactions. This output, then, is the result of the particular phoneme sequence chosen, with the inherent (linguistic) durations of each phoneme, as well as the influences of phonetic, semantic, and syntactic context in which the phonemes occur. The second level postulated by Bonnot is the motor execution level, or "clock time" (Kent, 1983) whose output is the result of the biomechanical properties of the muscles and other structures of the speech system.

This study was designed to exploit the idea that Central Nervous System pathologies can be used to test the hypothesis of separate motor planning and execution levels for speech (that there is a phonological—or linguistic—stage and a motor execution stage). Our specific hypothesis was that, in the case of CNS pathology at the final stages of motor realization, we would find a global slowing of speech—equivalent lengthening both of vowel and consonant segments across an utterance—whereas with perturbations at higher levels we would find sequencing problems and disturbances of timing patterns. We have begun our investigations with a study of ataxic dysarthria, in which there is cerebellar disease or damage. (The cerebellum (see, e.g., Ito, 1984) is thought to be important in initiating and terminating movements, as well as coordinating complex movements). We expect this disorder to lead to execution-level, rather than planning-level problems. In terms of compensatory shortening and terminal lengthening, the persistence or absence of these effects would lend support to placing their origin at the planning or execution level.

Acoustic investigations of speech timing in this dysarthria report lengthening of segments (Kent & Netsell, 1975; Kent, Netsell, & Abbs, 1979; Kent & Rosenbek, 1982) and increased mean syllable durations (Linebaugh & Wolfe, 1984) compared to normal speakers. There is some disagreement about compensatory shortening by these speakers: Kent et al. (1979) reported that their ataxic subjects did not demonstrate the normally occurring phenomenon of compensatory shortening in stem+suffix utterances of increasing numbers of syllables, whereas Mermelstein (1989) and Bell-Berti, Lazowski, and Mermelstein (1989) reported that compensatory shortening tended to be preserved in their ataxic speakers. There is agreement among several investigators that ataxic speakers do not demonstrate the normal phenomenon of terminal lengthening (Bell-Berti et al., 1989; Mermelstein, 1989; Yorkston, Beukelman, Minifie, & Sapir, 1984), although none of these studies has provided detailed descriptions of the changes in segmental timing across utterances.

Methods

Language. We chose to study French because it is a language with bound stress: That is, the placement of stress on a syllable is determined with reference to the word, and the position of stress identifies the word as a phonological unit (see, e.g., Lehiste, 1976). The situation is different in English, a language with free stress, in which the placement of stress is not predetermined by the morphological, lexical, or syntactic rules of a language. By studying French we could avoid the potentially substantial effects that "free stress." ·

Subjects. The subjects for this study were adult native speakers of Parisian French. The two dysarthric subjects suffered from Friedreich's Ataxia; they were located with the help of l'Association Française de l'Ataxie de Friedreich. One of these subjects was able to walk with only modest assistance; the second was unable to transfer between chairs without assistance. The three control subjects were speakers without any known neurological, speech, or auditory pathology.

Speech materials and recordings. A minimal set of twelve French nonsense words was created; each word was placed in two carrier phrases, one "short" and one "long," resulting in 24 sentences. The words were designed to vary systematically in the number of word-initial consonants as well as in vowel quality (Table 1). The sentences were placed in a list in random order and presented to each subject, to be read aloud.

The recordings were made in a sound-treated booth. The three control subjects and one dysarthric subject each repeated the list ten times (producing 240 sentences each). Because of obvious subject fatigue, the recording session with the second dysarthric was curtailed after he had produced only seven repetitions of the list, or 168 sentences. The subjects had no difficulty with the task: they produced each sentence on a single breath, without pauses. Subjects were allowed to self-correct their productions. The dysarthric subjects did have some articulation errors, especially a dentalization of /s/ that approached [θ] in quality. Except for the occasional production of a "wrong" sentence (e.g., one of the others), all sentences were measured (401 of 408 recorded for the dysarthric speakers; 716 of 720 recorded for the control speakers). The acoustic speech signal was recorded on a Revox A77 tape recorder and the speech waveform was simultaneously written, via a Gould ES1000 recorder, to paper traces, from which all measurements were made.

Table 1. *Experimental utterances: Orthographic and phonemic representations.*

tide	[tidə]	tude	[tydə]	toude	[tudə]
side	[sidə]	sude	[sydə]	soude	[sudə]
stide	[stidə]	stude	[stydə]	stoude	[studə]
stride	[stʁidə]	strude	[stʁydə]	stroude	[stʁudə]

Carrier phrases:
short: C'est la _________ .
long: C'est la triste _________ .

Measurements. The overall duration of each repetition was measured, as well as the duration of each of the acoustically identifiable segments of "C'est la" (/s/ and /ela/—the /e/, /l/, and /a/ could not be segmented) and of the initial consonant sequence of each target word, as well as its vowel, /d/, and terminal schwa. Means and standard deviations were calculated for each measure for each subject group, control and dysarthric, and t-tests performed to identify differences between durations in the speech of the control and the dysarthric subjects as well as between durations in short and long carrier-phrase utterances within each subject group.

Results

The first point we wish to make is that, in agreement with reports in the literature, the mean durations of the dysarthric speakers' utterances are significantly greater than those of the control subjects for all 24 sentence comparisons (Figure 1). Furthermore, this "slowness" does not result from pauses within the sentences. The question, then, is where this difference in duration resides: that is, are there durational increases across the sentence, and do sentence-length timing patterns survive the temporal disturbances of the dysarthria? Examining the durations of individual segments, we find that the utterance-initial /s/ and /ela/ segments are significantly longer in the speech of the dysarthric subjects (Figure 3). But do these increases override such sentence-length effects as compensatory shortening? In this data set, there was a tendency for compensatory shortening to occur, although it was not significant. The result is stronger, however, in a larger data set collected from these five subjects and one additional dysarthric subject. This larger set includes the data reported here as well as data from a parallel set that used the same carrier phrases target words that are French lexical items. In the pooled data, compensatory shortening effects were found for the control and dysarthric subjects for both the utterance-initial /s/ and the /ela/ of the carrier phrase (Table 2). That is, even though the dysarthric speakers produce initial /s/ and /ela/ acoustic segments with significantly greater durations than do the control speakers, they maintain the durational relations of these segments across phrases of different lengths.

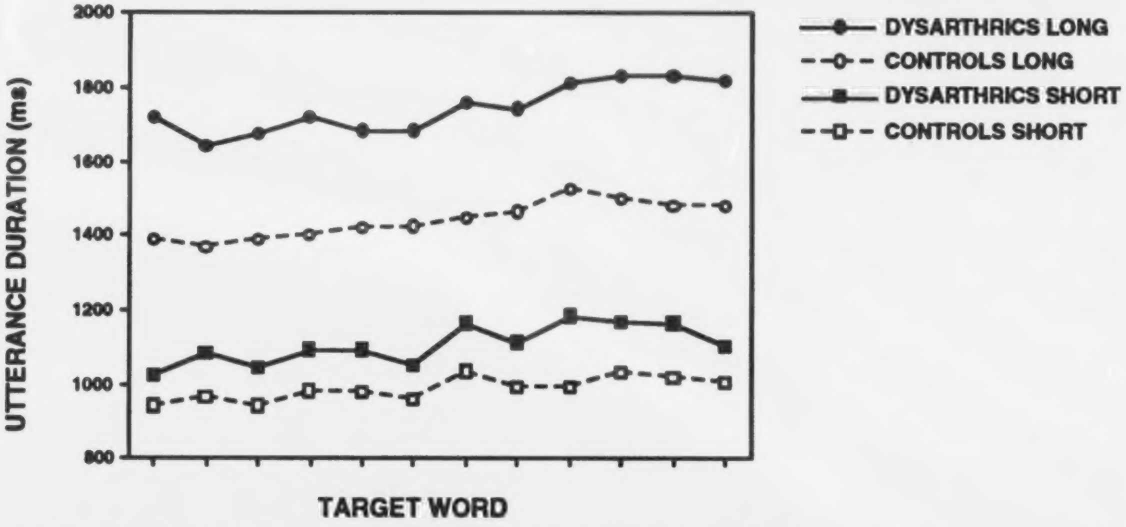


Figure 1. **Mean durations of each sentence for dysarthric subjects (filled symbols) and control subjects (open symbols), for short carrier phrase utterances (square symbols) and long carrier phrase utterances (round symbols), for each sentence. Target-word initial consonants and vowel are indicated along the abscissa. All differences are significant ($p < .05$).**

Table 2. *Compensatory shortening effects: Durational differences between short and long utterances (in ms.)*

	/s/	/ela/
Control n=712 sentences	8.5**	16.6**
Dysarthric n=514 sentences	7.8*	26.1**

*$p < 0.05$
**$p < 0.0001$

Returning to the examination of segment-duration patterns within an utterance (Figure 3), we must look at the durations of the target-word segments. Although the dysarthric speakers tend to have longer target-word initial consonant sequences than the normals speakers the difference is significant in only six of the 24 sentence comparisons, three short sentences and three long sentences. We also find a failure of the dysarthric subjects to produce target-word vowels and /d/ with greater duration than the control subjects; indeed, in the three significantly different comparisons for each of these segments, the durations are greater for the control subjects. Finally, the duration of utterance-final schwa is always significantly different for these subject groups, but the durations are always greater for the normal than for the dysarthric speakers—in fact, there is no overlap in the duration ranges of the groups' [ə] durations (Figure 2).

We have looked at the timing pattern within the sentences of the two subject groups in two ways. First, we have plotted the mean duration of each of the measured intervals for the short and long sentences for both groups. Figure 3 clearly reflects the greater duration early in the sentences of the dysarthric subjects, the similarity of target-word segment durations other than schwa, and the greater schwa duration in the sentences of the control subjects. A second way of looking at these utterance-length patterns is to examine ratios of the durations of the segments produced by the dysarthric and control subjects. We have done this by dividing the durations of the dysarthric subjects segments by those of the controls (Figure 4). We see that the ratios are greater than 1.0 for the early segments (/s/ and /ela/), approximately equal for the medial segments (/C_n/, V, and /d/), and less than 1.0 for the final schwa (were the dysarthric speakers had shorter durations than the normals). That is, the *relative* durations of the dysarthric subjects' segments are shorter later in the sentences.

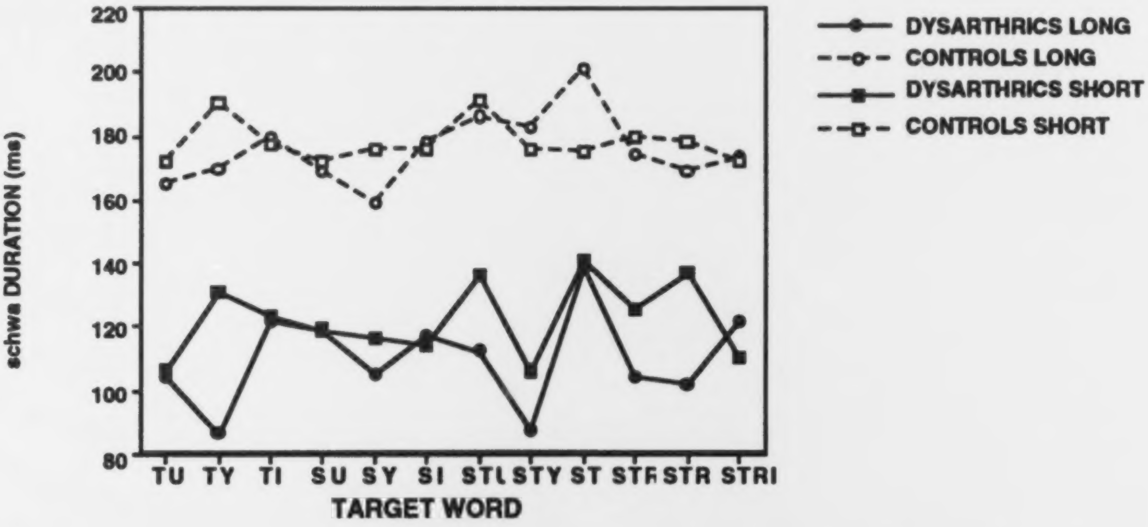


Figure 2. Mean durations of target-word final schwa for each sentence for dysarthric subjects (filled symbols) and normal subjects (open symbols), for short sentences (square symbols) and long sentences (round symbols). All differences are significant ($p < .05$), but control subjects' vowels are longer than those of dysarthric subjects.

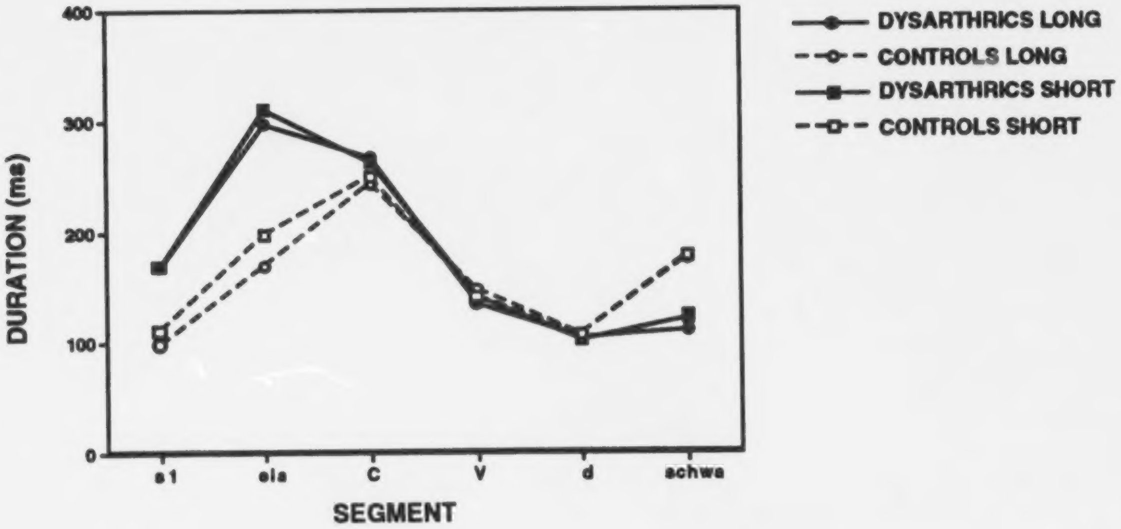


Figure 3. **Mean durations, pooled across short sentences and long sentences, of each measured interval: carrier-phrase initial /s/ and /ela/, and target-word initial consonants ("C"), vowel ("V"), /d/, and final schwa. Data for dysarthric subjects are plotted with filled symbols and connected with solid lines; those for control subjects are plotted with open symbols and connected with dashed lines; data for short sentences are plotted with square symbols, those for long sentences with round symbols.**

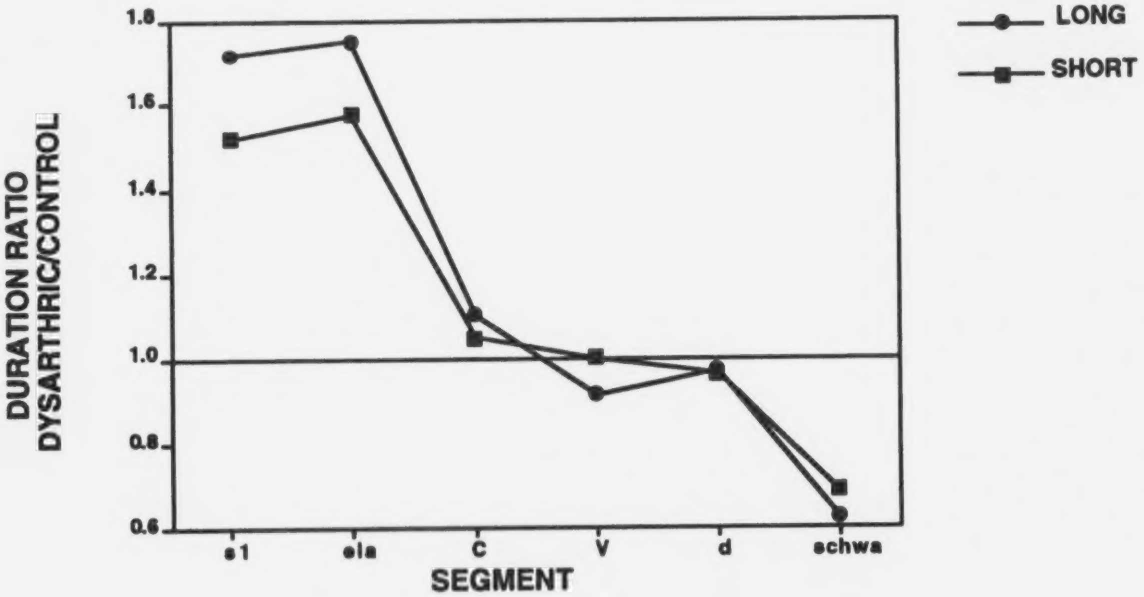


Figure 4. **Ratios of dysarthric subjects' to normal subjects' segment durations for short sentences (squares) and long sentences (circles).**

CONCLUSIONS

These data suggest to us that at least some aspects of the timing patterns of fluent speech arise in the cerebellum. First, since we have found a general decrease in speaking rate in our dysarthric speakers, with overall utterance duration being greater for these subjects than for the control subjects, we conclude that the cerebellum influences speaking rate. Second, since compensatory shortening was observed in the speech of these cerebellar dysarthric speakers, we suggest that the linguistic/motor planning level, at which notional timing arises, was preserved Third, the slowing of speech in these dysarthric speakers is not simply the result of a global slowing of speech—it is not analogous to playing back a tape recording at a slower speed. Rather, the timing patterns in these dysarthric subjects' speech supports the idea that the cerebellum has a "setting" function that is generally applied to units of, perhaps, breath group size, and that this function has been lost. Fourth, since the durations of the control subjects segments equal and then become longer than those of the dysarthric subjects as they proceed through the sentences, we suggest that terminal lengthening is an execution-level, and not a planning level, effect. Further study is, of course, needed to test these ideas, and we are currently undertaking such a project .

REFERENCES

Bell-Berti, F., Lazowski, G., & Mermelstein, P. (1989). Temporal characteristics of speech: implications of ataxic dysarthric speech timing. Paper presented at the annual meeting of the American Speech-Language-Hearing Association, St. Louis, MO 11/19/89.

Bonnot, J.-F. (1989). Timing intrinsèque et timing extrinsèque: le temps est-il une variable contrôlée? *Journal d'Acoustique*, 2, 287-296.

Darley, F., Aronson, A. E., & Brown, J. R. (1975). *Motor Speech Disorders*. Philadelphia: W. B. Saunders Company.

Ito, M. (1984). *The cerebellum and neural control*. New York: Raven Press.

Kent, R. D. (1983). The segmental organization of speech. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 57-89). New York: Springer-Verlag, .

Kent, R. D., & Netsell, R., (1975). A case study of an ataxic dysarthric: Cineradiographic and spectrographic observations. *Journal of Speech and Hearing Disorders*, *40*, 115-134.

Kent, R. D., Netsell, R., & Abbs, J. (1979). Acoustic characteristics of dysarthria associated with cerebellar disease. *Journal of Speech and Hearing Research*, *22*, 627-648.

Kent, R. D., & Rosenbek, J. C. (1982). Prosodic disturbance and neurologic lesion. *Brain and Language*, *15*, 259-291.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, *59*, 1208-1221.

Klatt, D. H., & Cooper, W. E. (1975). Perception of segment durations in sentence contexts. In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech production*. Heidelberg: Springer Verlag.

Lehiste, I. (1976). Suprasegmental features of speech. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225-239). New York: Academic Press.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm*, *21*, 1-59.

Linebaugh, C. W., & Wolfe, V.E. (1984). Relationships between articulation rate, intelligibility, and naturalness in spastic and ataxic speakers. In M. McNeil, J. C. Rosenbek, & A. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management*. San Diego: College-Hill.

Mermelstein, P. (1989). *Temporal characteristics of the speech of normal and Friedreich's ataxic speakers of French: A comparative study*. Unpublished M. A. Thesis, St. John's University.

Nakatani, L. H., & Schaffer, J. A. (1978). Hearing words without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, *63*, 234-245.

O'Shaugnessy, D. (1974). Consonant durations in clusters. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *22*, 282-295.

Tatham, M. A. A. (1970). A speech production model for synthesis-by-rule. *Ohio State University Working Papers in Linguistics*, *6*. (Cited in R. D. Kent, The segmental organization of speech. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 57-89). New York: Springer-Verlag. (1983))

Yorkston, K. M., Beukelman, D. R., Minifie, F. D., & Sapir, S. (1984). Assessment of stress patterning. In M. McNeil, J. C. Rosenbek, & A. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management*. San Diego: College-Hill.

FOOTNOTES

[*] In H. Peters, W. Houlstijn, & C. W. Starkweather (Eds.), *Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering* (1990).

[†] Also St. John's University, Jamaica, New York.

[††] NSERM,Laboratoire de Physiologie et Pathologie Cérébrales, Paris, France.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 93-110

Implications of the Concepts Underlying Task-dynamic Modeling on Kinematic Studies of Stuttering*

Peter J. Alfonso†

The assigned task is to demonstrate how the concepts underlying task-dynamic modeling can improve our understanding of speech motor organization in those who stutter. The discussion focuses on the principles underlying, and the results of experiments that have centered upon, certain organizational principles of skill motor behavior. Based on the results of experiments of this type, the following assumptions are drawn: (1) the stutterers' speech motor system exhibits generalized spatio-temporal dysfunction and (2) the dysfunction occurs at a level of motor control that is generally responsible for spatio-temporal organization of all skilled motor activities. This implies that the disorder of stuttering is associated with, in addition to the linguistic factors that are encoded in speech movements and the behavioral influences which might mediate them, irregularities at a rudimentary and pre-linguistic level of motor control. An example of how these assumptions can be tested by task-dynamic modeling is given.

This paper will discuss how the concepts underlying task-dynamic modeling can help us understand better speech motor control in those who stutter. To this end, I discuss the results of a select body of research that explores the motor organization of speech by stutterers, and conclude that the research suggests that perceptually fluent speech gestures produced by stutterers do not follow certain general principles of motor organization in human behavior. The discussion will be limited to stutterers' speech that is perceived to be fluent, partly because of time constraints, but more importantly because it is not clear whether stutterers' speech movements during their perceptually fluent speech, and by inference their control of the speech mechanism, are different, in some way, from nonstutterers. First some experimenters have concluded that stutterers' perceptually fluent speech is not different than control subjects' fluent speech, while other experimenters think that it is (see, for example, Bloodstein, 1987, Chapter 1; Van Riper, 1982, Chapter 16).

Although there are a number of reasons why the results of experiments on stutterers' perceptually fluent speech are in conflict, the conflicting results are undoubtedly confounded by the general lack of adequate definitions and criteria for distinguishing among certain essential characteristics of stuttering, for example, involuntary dysfluency and perceptual fluency. This issue is discussed recently in considerable detail in "Research Needs in Stuttering: Roadblocks and Future Directions" (Cooper, 1990) and need not be elaborated here.

A second reason for the conflict in the results is that most of these experiments are based on perceptual and speech acoustic data, and few include kinematic and neuromuscular data in parallel with perceptual and acoustic, although it is generally the case that perceptual and acoustic data in the absence of simultaneously gathered physiological data are not sufficient to answer questions about speech motor control. This is a significant problem with regard to the fluency issue because

---

Preparation of this paper supported in part by NIH Grant DC-00121 and BRS Grant RR-05596 awarded to Haskins Laboratories. Final version of this paper was completed during the author's tenure as Fulbright Research Scholar, The University of Nijmegen, The Netherlands. Grateful acknowledgment is extended to Dr. Wouter Hulstijn for his careful and constructive review of the final draft of this paper.

while a segment of stutterer's speech may appear normal or "fluent" at the perceptual level, it may appear abnormal or "dysfluent" at the acoustic, and/or movement and muscular levels (Alfonso, Watson, & Baer, 1984; Baer & Alfonso, 1984; Shapiro, 1980).

A third reason for the conflict in the results of the reported literature is that the routinely posed form of the question, "Is stutterers' fluent speech similar to the fluent speech of adults who do not stutter," is too broad. It is likely that certain aspects of speech do not differentiate the groups while other aspects do differentiate the groups. Using respiratory comparisons as an example, Baken, McManus, and Cavallo (1983) found no differences between stutterers and control subjects in chest wall preposturing maneuvers immediately preceding fluent speech. Yet, significant group differences in subglottic pressure (e.g., Lewis, 1975; Peters & Boves, 1988), flow rates (e.g., Hutchinson, 1975), and lung volume charge and deflation for speech (e.g., Watson & Alfonso, 1987; Story, 1990) have been observed. Thus, it might be that stutterers and normal speakers perform certain speech respiratory gestures in a similar fashion, for example, respiratory preposturing, but that other aspects of speech respiration, for example, the magnitude of the inspiratory charge, are performed differentially. With respect to task-dynamic modeling, the concepts underlying the current version of the model would be appropriate for addressing specific questions about supralaryngeal articulation, but not speech production in general.

A fourth reason for the conflict in the results of these types of experiments, and the one most pertinent to the main focus of this paper, is that group comparisons are frequently based on physiological data that reflect relatively variant phonetic level speech gestures. Group data based on phonetic level contrasts are inherently unstable since spatial and temporal control of the speech structures to mark phonetic distinctions varies as a function of phonetic context, stress and rate, dialect, and individual speaker preferences. Rather, we need to make group comparisons based on relatively stable spatial and temporal characteristics of normal speech dynamics, for example, those that best reflect organizational principles of speech motor control (Alfonso, Story, & Watson, 1987a,b; Alfonso, Watson, & Seider, 1986; Caruso, Abbs, & Gracco, 1988).

Dynamic parameters that are relatively invariant across multiple productions of an utterance are thought to be good representatives of speech motor control parameters (e.g., Gracco & Abbs, 1986) and form the bases for the notion of coordinative structures, from which task-dynamic, and other models of speech production, are empirically formed. Implicit in this notion is that the individual articulator components comprising a multiarticulate gesture are functionally linked into coordinative systems. These relatively stable spatial and temporal characteristics could also serve as criteria for distinguishing between a normal or fluent motor system and an abnormal or dysfluent motor system. Examples of stable dynamic parameters that could serve as criteria are; (1) displacement/velocity ratios and peak velocity profiles, (2) motor equivalence covariability, and (3) inter-articulator relative-timing and sequence patterns. As an example of the first criterion, the kinematics associated with the individual articulator components that comprise, for example, an alveolar closure gesture, the tongue tip and tongue blade, and the jaw in this example, as well as the alveolar gesture itself, that is, the additive contribution of the anterior tongue and the jaw, should demonstrate smooth and single peaked velocity profiles. On the other hand, group comparisons based solely on the amplitudes of the individual articulator displacements, particularly in absence of other related movement data, should be treated with some degree of caution since speakers achieve alveolar closure gestures with distinctly different contributions of individual tongue and individual jaw displacement, thus reflecting idiosyncratic speaker preference rather than a general speech motor characteristic common to all speakers within the group.

A second criterion would focus upon the organization of the relative tongue-jaw displacement, that is, to determine whether the gesture demonstrates motor equivalence covariability, which reflects the way in which the motor system reorganizes itself en route to a task goal when faced with either internal or external perturbations. In normal fluent motor systems, when one structure in coordinated movement is perturbed, a response will be observed in all structures to which the perturbed structure is linked by a common task. The two ways in which motor equivalence is generally assessed reflects the motor systems response to internal or external perturbations. With respect to the motor system response to internal perturbations, in repeated-trial tasks, the magnitude of the variability associated with the displacement of the vocal-tract gesture, the alveolar closure in this example, should not be greater than either of the magnitudes of the variabilities associated with the individual articulator displacements comprising the gesture, the tongue and the jaw in this example. A second way that is often used to test motor equivalence is the traditional perturbation paradigm, where an unexpected external force is applied to a speech structure. In the example of the alveolar closure gesture, an external perturbation applied to the jaw should elicit a compensatory response to the anterior portion of the tongue but not in the lower lip.

The third criterion focuses primarily on certain temporal characteristics of speech, specifically, on inter-articulator relative-timing and sequencing, since in normal systems consistent latencies among and ordering of articulator movements are observed in specific phonetic contexts.

Review of Research Findings

There are relatively few, but all recent studies, that have examined these types of parameters in stutterers' perceptually fluent speech. Of course, we need to ascertain what the values of these parameters are before we can implement task-dynamic modeling, or any other physiological based model for that matter, to improve our understanding of stutterers' speech motor control. In what follows, I will first discuss the results of some of these experiments, primarily to demonstrate examples of data that could be used as certain dynamic parameters of the model. Because of the significant differences in various aspects of these experiments, for example, in the methodology, speech tasks such as phonetic context and rate, numbers and severity of stutterers and their therapeutic histories, and to a lesser extent the in\strumentation, a comparison of the results is not, in all cases, straightforward. Because of this, I will begin each topic by first discussing the differences in the experiments reported by Caruso et al. (1988) and Alfonso et al. (1987a,b); Alfonso, Watson, and Baer (1987). I do this primarily because they were among the first experiments specifically designed to examine the motor control parameters referred to above, and because they represent, in some cases, the extreme differences in the results, they serve as a good point of departure in regards to the ubiquitous intersubject variability issue that is generally observed in stutterers, and in some cases, control subjects' data. Lastly, I will demonstrate some of our early attempts to use task-dynamic modeling by modifying certain dynamic parameters of the model to conform with data obtained from our experiments that investigated adult stutterers' tongue and jaw movements in perceptually fluent speech (Alfonso et al., 1987a,b,).

Turning to the first criterion, where single-peaked velocity profiles serve as an index of coordinated multiarticulator systems, Figure 1 shows velocity profiles for upper lip, lower lip, jaw, and combined movements for the closing gesture in the first /p/ in 'sapapple' taken from Caruso et al. (1988). In this experiment, the movements of the lips and jaw were transduced by means of strain-gauges attached to light-weight head-mounted cantilever beams. The trajectories on the left are from a normal speaker and the trajectories on the right are from a stutterer. Note that the velocity profiles for each of the articulators and the combined signal, which represents a labial closing gesture here, is characterized by smooth and singled-peaked velocity profiles indicating that stutterers are capable of organizing, in the spatial and temporal domains, the movements of the individual articulators to yield well-coordinated labial closure and release gestures.

Figure 2 shows a different view of peak velocity and is taken from Alfonso et al. (1987a). The figure shows peak vertical displacement amplitude by derived peak velocity for lower lip, tongue blade, and jaw movements during initial /s/ closure in multiple productions of /esese/ for a single severe stutterer. Articulator displacement was derived by the x-ray microbeam pellet tracking system at the University of Tokyo. Small lead-pellets were glued to the lips, tongue, and jaw for this purpose. For the sake of clarity, the data points associated with the movement of each structure have been enclosed within solid lines. Note that the same general relationship between displacement amplitude and peak velocity previously reported for limb movements and lip-jaw movements holds for these data as well (Kelso, V.-Bateson, Saltzman, & Kay, 1985). That is, peak velocity varies directly with displacement amplitude; for example, peak velocity increases as displacement amplitude increases. The correlations for peak displacement amplitude and peak velocity for stutterers and controls in this experiment varied from .89 to .95. Figure 3, taken from De Nil and Abbs (1989a), shows similar results for lip and jaw movements, which were transduced by a strain-gauge system similar to that in the Caruso experiment. The figure shows velocity/displacement ratios across variable speech rates and demonstrates that both stutterers and control subjects scale peak velocity as a function of the intended displacement amplitude.

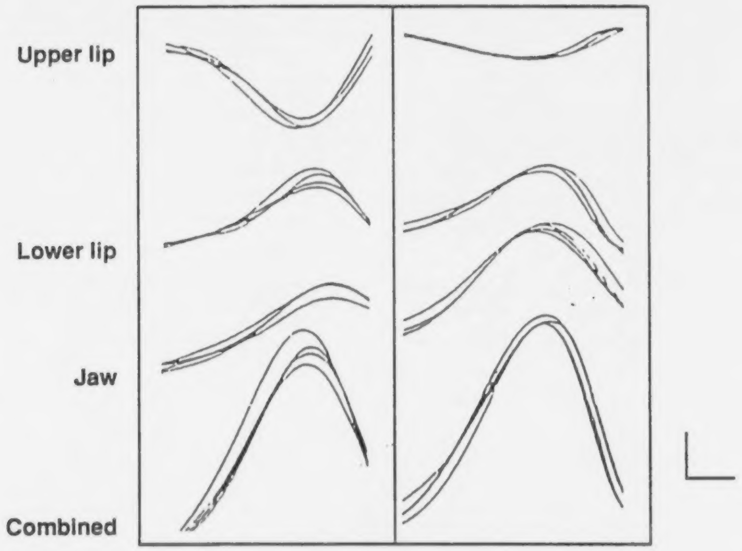


Figure 1. **Taken from Caruso, Abbs, and Gracco (1988). Superimposed velocity profiles for upper lip, lower lip, jaw, and combined closing movements (4 trials) for the first "p" in "sapapple" for 1 normal speaker (*left*) and 1 stutterer (*right*). Individual movements and the combined signal for both the stutterer and the normal speaker manifested single-peaked velocity patterns. Horizontal and vertical calibrations represent 20 ms and 70 mm-s, respectively.**

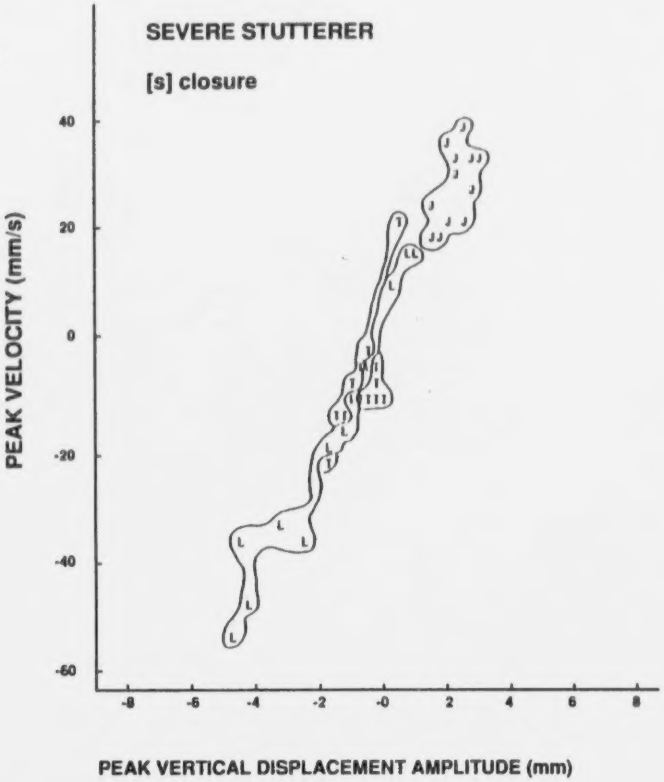


Figure 2. **Taken from Alfonso, Story, and Watson (1987a). Peak vertical displacement amplitude by derived peak velocity for lower lip (L), tongue blade (T), and jaw (J).**

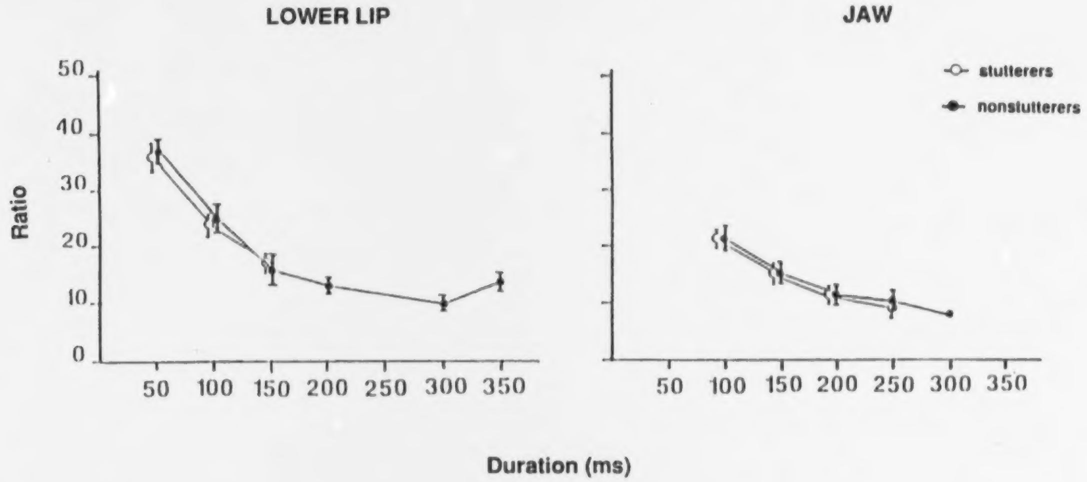


Figure 3. **Taken from De Nil and Abbs (1989a). Ratio of peak velocity/displacement by closure duration.**

All of these data support the notion that certain displacement and velocity characteristics of stutterers' perceptually fluent speech, particularly with respect to individual articulator movements, are not distinguishable from the same characteristics of non-stutterers' fluent speech. However, intersubject variability is very high, particularly for stutterers, and seems to be related primarily to stuttering severity, therapy influence, and speech rate. I will return to the influence of these variables on stutterers' speech movements latter.

The second criterion centers on the demonstration of motor equivalence covariability. The coefficient of variation (CV), the ratio of the standard deviation and the mean, is often used here because it permits a comparison of displacement variability among different speech structures despite differences in absolute displacement for each of the structures. Table 1 is taken from the Caruso et al. (1988) experiment and shows the CV's for the closing gesture associated with the first /p/ in "sapapple." Recall that a well-coordinated multi-articulate system would demonstrate less variability for the vocal-tract gesture, the combined signal here, relative to the variability associated with each of the articulator members of the gesture, the upper and lower lip, and the jaw in this example. The table shows that the CV for the combined signal is smaller or equivalent to any of the CV's for the individual articulators for both the stutterers and the controls and suggests that both stutterers and controls demonstrate comparable levels of motor equivalence covariability.

Table 1. *Taken from Caruso, Abbs, and Gracco (1988). Coefficients of variation of the peak closing displacement for each speech structure and the combined signal. n=number of movements; UL=upper lip movement; LL=lower lip movement; J=jaw movements; C=combined signal.*

CONTROLS

	N1 n=176	N2 n=173	N3 n=175	N4 n=168	N5 n=71	N6 n=119
UL	0.204	0.303	0.149	0.125	0.095	0.251
LL	0.160	0.159	0.116	0.181	0.103	0.210
J	0.255	0.154	0.214	0.167	0.266	0.262
C	0.136	0.138	0.113	0.104	0.092	0.186

STUTTERERS

	S1 n=153	S2 n=172	S3 n=175	S4 n=124	S5 n=137	S6 n=172
UL	0.113	0.185	0.129	0.259	0.327	0.225
LL	0.067	0.123	0.149	0.132	0.078	0.124
J	0.197	0.172	0.114	0.206	0.139	0.138
C	0.069	0.123	0.090	0.105	0.083	0.086

Figure 4 shows a different result. These data are taken from the Alfonso et al. (1987a) experiment and represent the weighted average CV's for closure and release alveolar gestures, namely the alveolar obstruents /t,s,n/. The control subject demonstrates motor equivalence. The magnitude and the relative patterns of the control subject's data for alveolar obstruent closure and release gestures are equivalent to the lip-jaw CV's for the first /p/ in sapapple by a large group of control subjects reported by Gracco and Abbs (1986). Notice, however, that the stutterers' data are quite different. The combined CV is greater than the CV associated with jaw movements, and the CV associated with tongue movements is the largest of all. These data do not support the idea that stutterers control functionally linked articulators in such a way that would result in invariant gestural goals.

What can account for the differences in the results observed in these two experiments? Some possibilities are rather obvious. For example, the subjects in the Caruso et al. (1988) experiment, which included five stutterers, repeated the word "sapapple" at a normal rate while the subjects in the Alfonso et al. (1987a) experiment, which included only two stutterers, produced the nonsense words /etete, esese, enene/ in a reaction-time task.

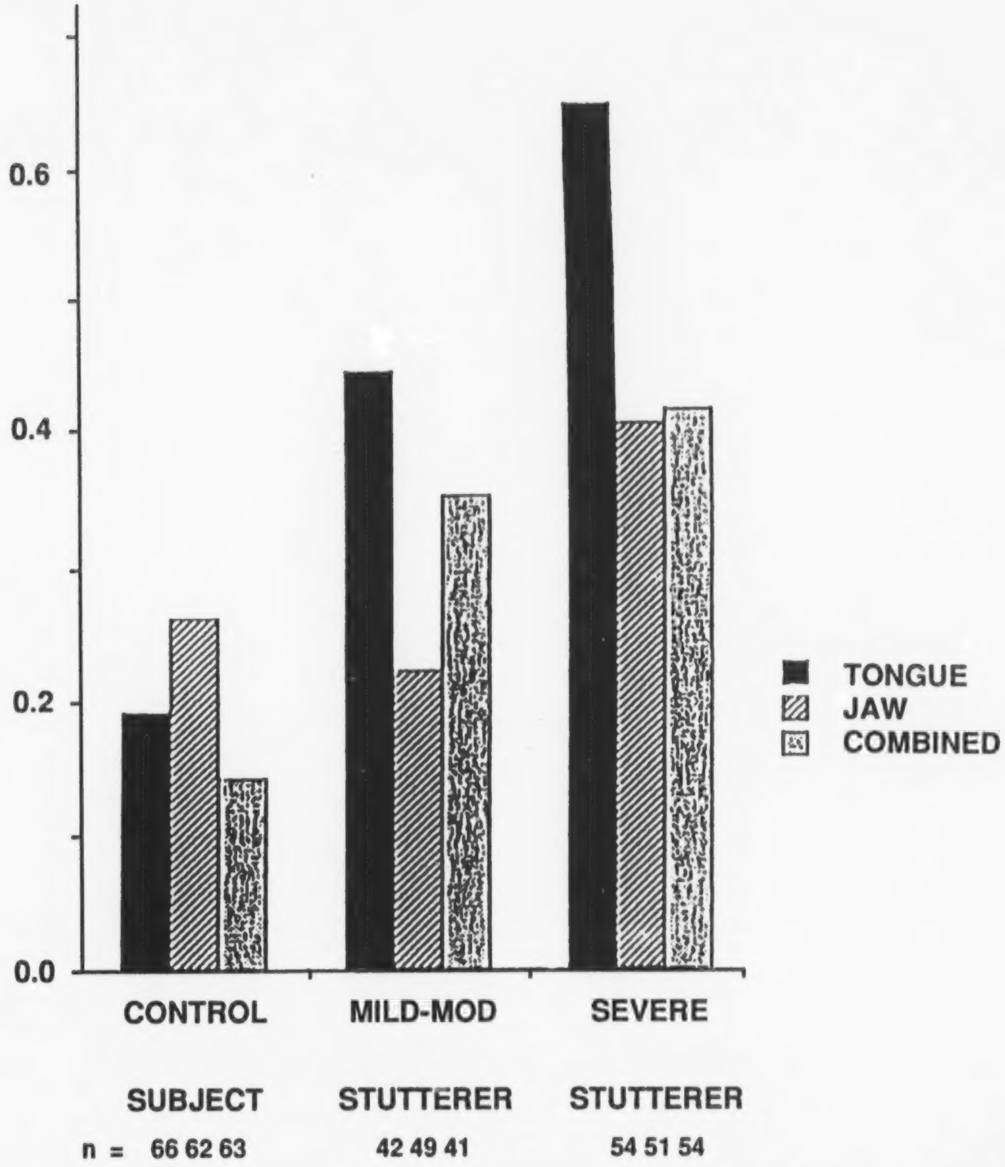


Figure 4. **Taken from Alfonso, Story, and Watson (1987a). Weighted average coefficients of variation for tongue blade, jaw, and combined tongue-jaw peak vertical displacement amplitude.**

The difference may be due to the relatively fast speech rate, the task, and the small number of stutterers in the Alfonso et al. (1987a) experiment, as well as other parameters that I will mention later. However, differences in the relative displacement of the individual articulators between the two groups of stutterers would seem to represent a significant contribution to the disparity in the results. Figure 5 shows that both controls and stutterers in the Caruso et al. (1988) experiment enlisted comparable amounts of displacement for both the lips and the jaw and the combined signal. While several of the group differences, including the combined signal for displacement and velocity, reached statistical significance (p. 443), the actual group differences were small. Recently, McClean, Kroll, and Loftus (in press), who also measured lip and jaw movements for /p/ closure in "sapapple," found no significant differences in the same displacement and velocity characteristics between controls and stutterers who had no history of speech therapy. However, significant group differences were found in stutterers who had recently completed therapy, although the differences were in articulator latency and in the duration of jaw movements and were most likely related to slower post-therapy speech rate. Contrary to these results, however, Figure 6 shows that the spatial organization of obstruent closure and release gestures (for simplicity, only /t/ and /s/ closure are shown here) is clearly different for the two stutterers compared to the one control subject in the Alfonso et al. (1987a) experiment. The figure shows the vertical displacement and associated peak velocity of a lead pellet attached to the tongue blade and demonstrates that the control subject, like all of the subjects in the Caruso experiment, achieves closure by complimentary activity of the individual articulators (the jaw and tongue blade) that comprise the gesture. However, the stutterers achieve closure primarily by jaw displacement with little, and occasionally paradoxical, tongue displacement. Thus, the control subject behaves like non-stuttering subjects in other experiments by demonstrating complementary movements of the speech structures while the stutterers behave in a qualitatively different fashion, notably by demonstrating paradoxical displacement of the tongue blade and jaw. The trial-to-trial variability in peak displacement amplitude and peak velocity, which is not shown in Figure 6, is much greater for the stutterers than the control subject. All of these differences were statistically significant. It could be that the stutterers' lack of movement covariability demonstrated in Figure 4 is related to their dominance of a single member of a supralaryngeal complex to achieve closure and release gestures.

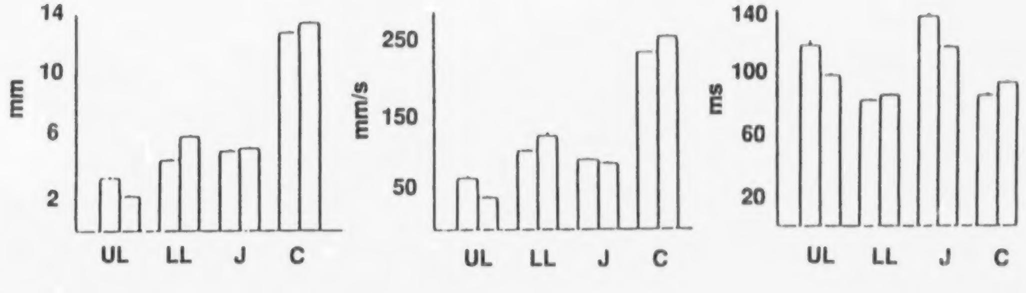


Figure 5. Taken from Caruso, Abbs, and Gracco (1988). Bar diagrams of UL, LL, J, and C movement characteristics (displacement (*left*), peak velocity (*middle*) and duration (*right*) for the 6 stutterers and 6 normal subjects. The vertical lines represent + 1 SEM. Normal subjects (open areas): stutterers (hatched areas).

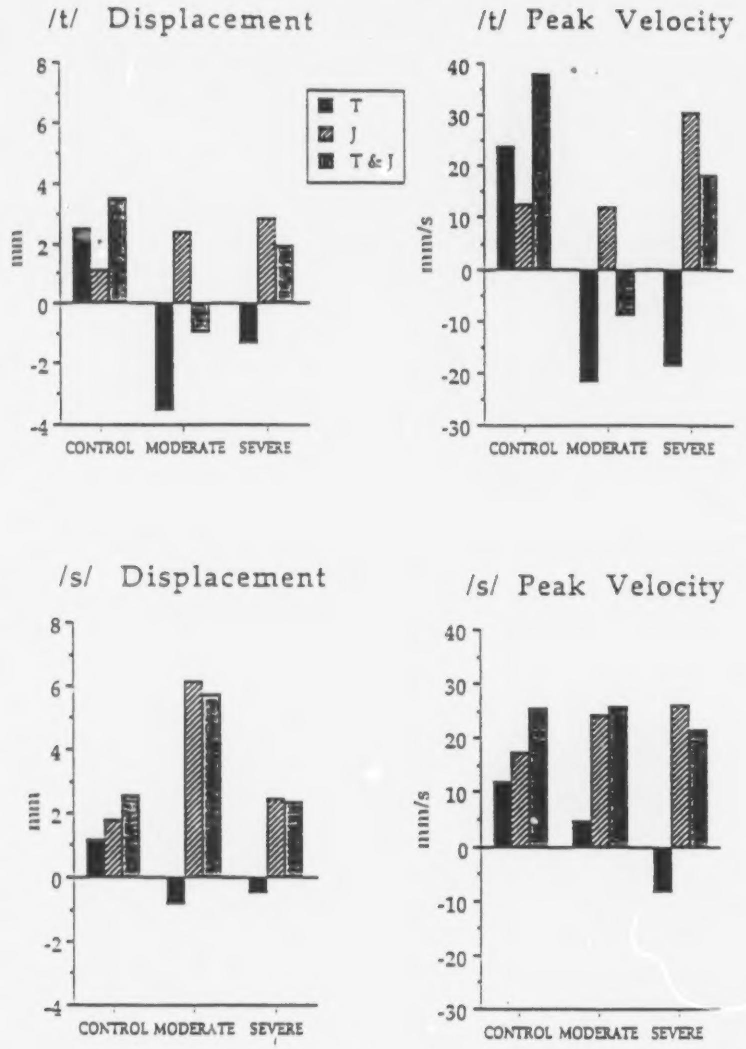


Figure 6. **Taken from Alfonso, Story, and Watson (1987a).**

More recent experiments shed some light on other possible influences on the differences observed between the Caruso and Alfonso experiments, namely therapy influences and speech rate control. A recently completed doctoral dissertation by Story (1990) examined the kinematics of stutterers' perceptually fluent speech as a function of therapeutic intervention and speech rate. Considering therapy influences first, Table 2 shows within-subject differences in mm between pre- and post-treatment conditions for UL, LL, J movements for /p/ closure (Table 2a) and release (Table 2b) in /pit/ and /pet/ produced in the carrier phrase "he see CVC again" for two controls and three stutterers. The stutterers were classified as severe pre-therapy and then mild post-therapy. Negative values indicate smaller post-therapy displacements relative to the corresponding pre-therapy values. Note that the pre- and post-treatment comparisons for both stutterers and controls reveal subject differences in the relative organization of the gestures. For example, control ES produced session 2 /p/ closure with much less J displacement and more UL displacement compared to his session 1 /p/ closure. Control DW, on the other hand, decreased J displacement but increased LL displacement. Most importantly, note that on average the combined displacement for closure and release gestures shows only about a one half mm pre- and post-therapy difference for the control subjects. On the other hand, the stutterers significantly decreased displacement amplitude post-therapy, note that the combined displacement pre- and post-therapy differences range from -1 to -4 mm. Similar results were obtained for /f/ closure and release in /fit/ and /fet/. All of the stutterers pre- and post-treatment difference values were statistically significant in all phonetic contexts, whereas a much smaller percentage of the control subjects' difference scores were significant.

Table 2a. *Taken from Story (1990). Displacement amplitude difference values (in mm) for /p/ closure in the words /pit/ and /pet/ at normal speech rate.*

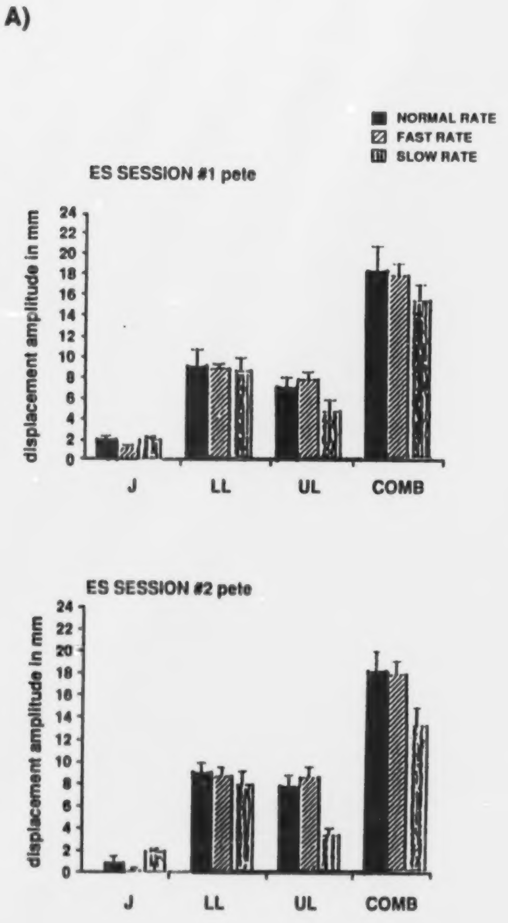
		CONTROLS		STUTTERERS		
		ES	DW	KH	PC	AB
J	/pit/	-.93	-.66	-.19	-1.19	-.84
	/pet/	-.78	-.48	-.31	-1.15	-.96
LL	/pi/	-.06	1.18	-.17	-.50	-1.58
	/pet/	-.40	1.19	-1.02	-.55	-1.54
UL	/pit/	.72	.25	-.11	-.64	-.03
	/pet/	.76	.02	.18	-.87	.16
Combined	/pit/	-.15	.78	-1.06	-2.32	-2.44
	/pet/	-.42	.73	-1.20	-2.57	-2.35

Table 2b. *Taken from Story (1990). Displacement amplitude difference values (in mm) for /p/ release in the words /pit/ and /pet/ at normal speech rate.*

		CONTROLS		STUTTERERS		
		ES	DW	KH	PC	AB
J	/pit/	.15	-.49	-.82	-2.48	-2.20
	/pet/	-.10	-.49	-.64	-2.75	-2.30
LL	/pit/	-.99	.52	-.20	-1.02	-1.31
	/pet/	-.93	1.14	-.35	-.76	-1.32
UL	/pit/	.22	-.51	-.18	-.37	-.29
	/pet/	.73	-1.27	-.12	-.75	
Combined	/pit/	-.62	-.48	-1.19	-3.86	-3.80
	/pet/	-.32	-.61	-1.10	-4.25	-3.62

Turning next to the influence of speech rate of stutterers displacement and velocity characteristics, Figure 7a, taken from Story (1990), shows /p/ closure displacement values for /pit/ and /pet/ for control ES at normal, fast, and slow speech rates. The acoustic duration of the phrase "he see pete (or pate) again" ranged from a mean of 750 ms at fast rates to a mean of 3500 ms at slow rates for the control subjects. Note that the session 1 (shown on top) and 2 (shown on the bottom) patterns are very similar. The normal and fast rates generate similar displacements while the slow rate generates relatively smaller displacements. Figure 7b shows the same conditions for stutterer KH. Phrase duration for the stuttering subjects ranged from a mean of 675 ms at fast rates to a mean of 5200 ms at slow rates. Note, first, that for all speech rates during the post-therapy condition (shown as session 2 at the bottom) displacements are less than their comparable pre-therapy displacements (shown as session 1 at the top) for all articulators and the combined signal, and second, that within the post-therapy condition, each articulator and the combined signal show less displacement at slow rates compared to normal and fast. This latter pattern is not consistently observed in the pre-therapy condition. The data for the remaining subjects are similar to those demonstrated in Figures 7a & 7b. The pre- and post-therapy comparison shown in Figure 7b could be the basis for the often cited observation made by Zimmerman (1980a,b) that stutterers' supralaryngeal displacements are slower and smaller in amplitude compared to normal speakers if the stutterers in his experiment were using therapy induced control strategies to maintain fluency.

Figure 8 is taken from De Nil and Abbs (1989a) and shows the relationship between jaw displacement (the left panel) and jaw peak velocity (the right panel) with /b/ closure duration in repeated utterances of /bae/. The differences between the groups become more distinct as speech rate decreases, that is, as closure durations increase. The group differences were significant at 150 ms and beyond. Once again, large intersubject variability for both groups was observed. The observation here that stutterers and controls increase displacement amplitude at slower rates is contrary to the Story (1990) observation of decreasing displacement at slow speech rates and may be attributable to a therapy effect on the Story (1990) stuttering subjects, and to speech utterance and/or the absolute rate differences between the experiments, since both the controls and stutterers decreased displacement amplitude at slow rates in the Story (1990) experiment. It should be noted that in the Story experiment the departure from normal to slow rates was much greater than the difference between normal and fast rates.



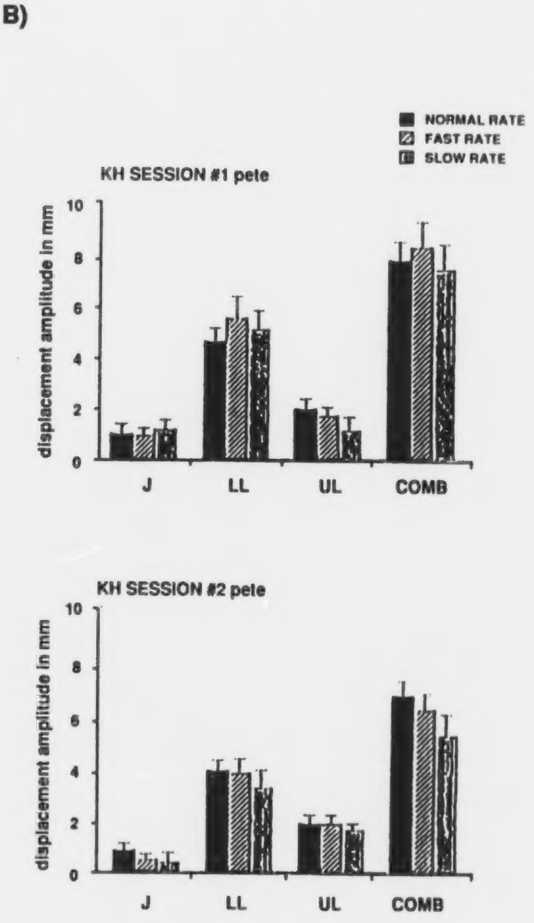


Figure 7a. **Taken from Story (1990). Control Subject.**

Figure 7b. **Taken from Story (1990. Stutterer.**

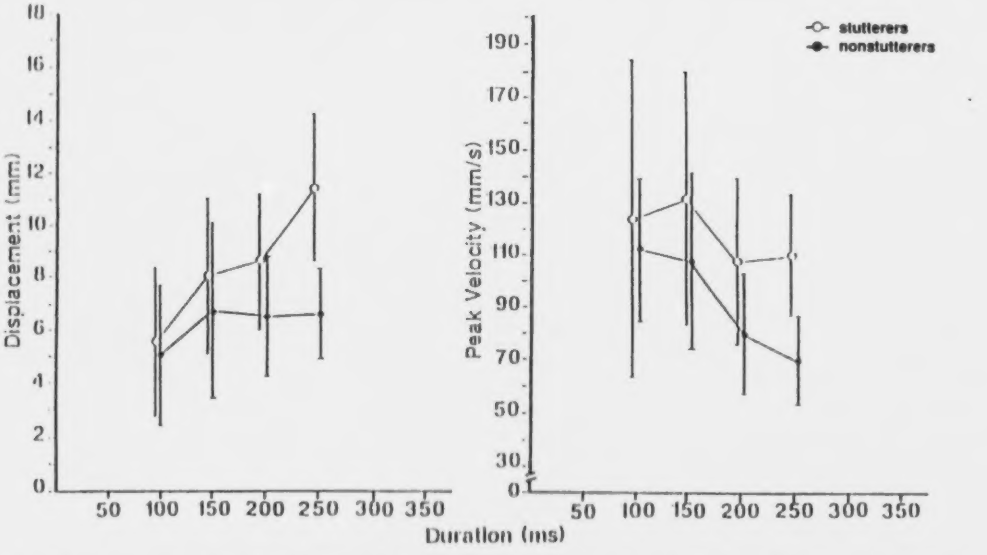


Figure 8. **Taken from De Nil and Abbs (1989a). Jaw displacement and peak velocity by closure duration.**

Turning to the last criterion, inter-articulator relative-timing and sequencing, the results, at first blush, appear less contradictory than those discussed above. Figure 9, taken from the Caruso et al. (1988) experiment, shows inter-articulator relative-timing and the temporal ordering of the upper lip, lower lip, and jaw for the first /p/ closure in "sapapple." Note that in the control subjects' data, shown on the left, the consistent order of movement is upper lip first, followed by the lower lip, and finally the jaw. This same sequence has been observed in larger groups of control subjects for the same task (e.g., Gracco & Abbs, 1986). However, only one of the six stutterers, shown on the right, showed this same sequencing pattern. In addition, the stutterers were found to be more variable in their sequencing patterns compared to the controls. The UL-LL-J temporal pattern in the "sapapple" context is most likely related to neural and biomechanical interactions and thus reflects differences in both neural control and biomechanical processes between stutterers and controls (Gracco, 1988).

Figure 10 shows the percent occurrence of context dependent tongue-jaw sequence patterns and demonstrates that similar results were observed for tongue and jaw movements in the Alfonso et al. (1987b) experiment. These data are pooled across the /t,s,n/ contexts and, like the Caruso data, represent temporal order relative to the peak in the velocity profiles associated with the displacement of each of the articulators. Figure 10 is of interest here because for the control subject different temporal patterns were observed for different phonetic contexts. For example, the tongue leads in /t/ closure but the jaw leads in /s/ and /n/ closure. For the stutterers, the tongue leads regardless of the phonetic context. The figure shows that the control subject demonstrates a consistent context dependant sequence in about 95 percent of the utterances while the stutterers showed the same sequence in about 60 percent of the utterances. As in the Caruso data, the stutterers' sequencing patterns were more variable relative to the control subject. Figure 11 shows the variability associated with tongue-jaw relative-timing. Here, the CV for interarticulator latency is plotted, which normalizes for intersubject differences in speech rate. The figure demonstrates that the two stutterers' show much more token-to-token variability in tongue-jaw relative-timing compared to the control subject, and supports the conclusions drawn from other experiments that stutterers demonstrate different temporal organization of the speech structures even during their perceptually fluent speech production. The group difference is statistically significant and is consistent with the relative displacement variability for these same subjects (see Figure 4) where the control subject shows the least variability and the severe stutterer shows the most.

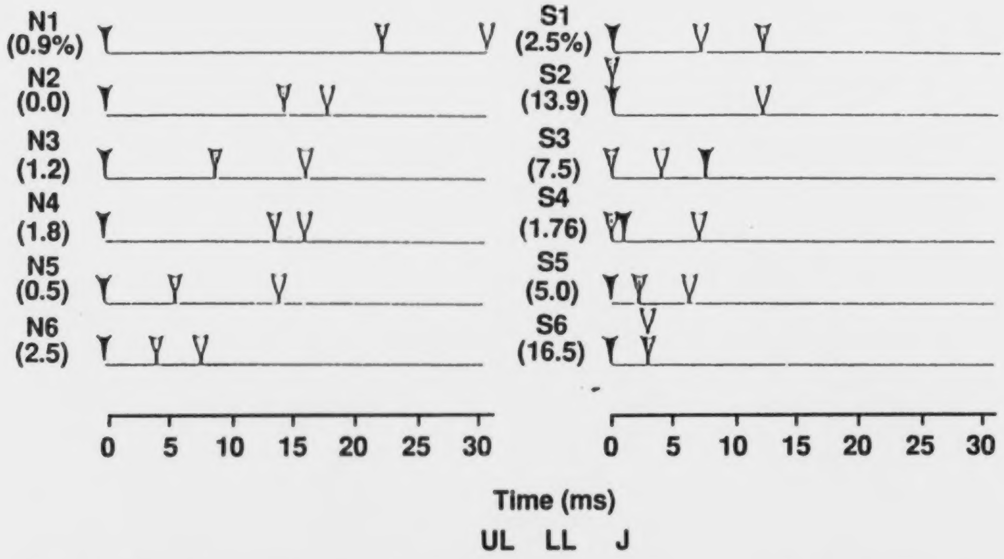


Figure 9. **Taken from Caruso, Abbs, and Gracco (1988). Schematic representations of the temporal ordering of UL, LL, and J peak velocities for each of 6 stutterers (S) and 6 normal (N) speakers. Percentage of sequencing reversals are shown in parentheses for each subject.**

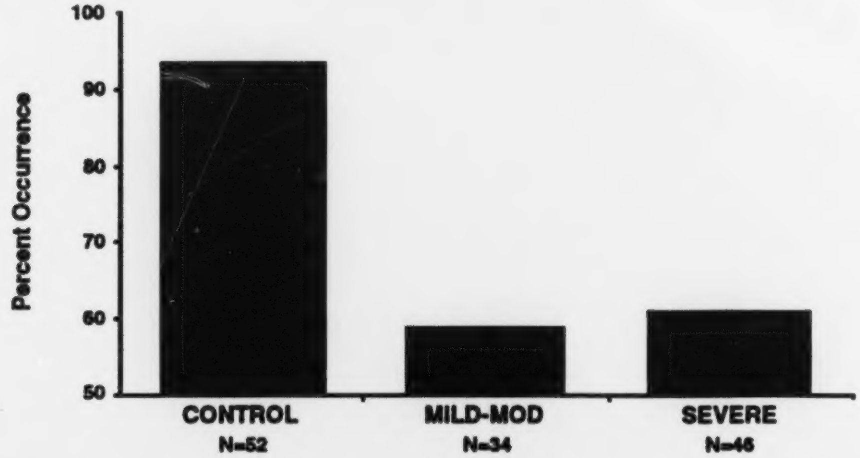


Figure 10. **Taken from Alfonso, Story, and Watson (1987a). Stutterers' tongue-jaw sequence compared to control by utterance type re velocity peaks.**

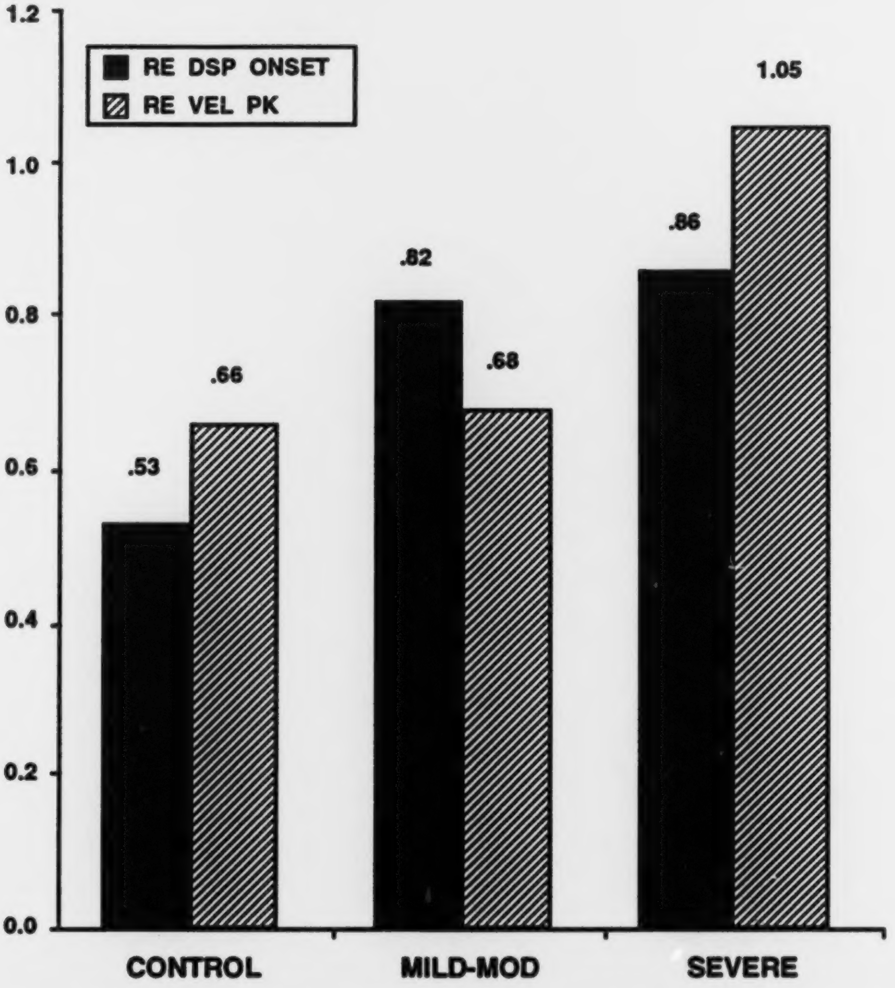


Figure 11. **Taken from Alfonso, Story, and Watson (1987a). Weighted average coefficients of variation for tongue-jaw relative time re displacement onset and peak velocity.**

More recent experiments have shown that interarticulator relative-timing and sequence patterns for stutterers can vary as a function of stuttering severity and program specific fluency-enhancing techniques, and in the case of both stutterers and control subjects, can also vary as a function of speaking rate and phonetic context. Figure 12, taken from Story and Alfonso (1989), shows an example of severity and therapy influences on lip and jaw relative-timing and sequence patterns during /p/ closure for perceptually fluent productions of /pit/ in "he see pete again." The data for two controls, shown on the left, represent two different sessions about six weeks apart and are consistent with the results of the Caruso et al. (1988) experiment with respect to both interarticulator relative-timing and sequence patterns. The stutterers data, shown on the right, are quite different. Recall that these stutterers were classified as severe pre-therapy and mild post-therapy. Considering interarticulator latencies first, note that pre-therapy latencies for stutterer AB (specifically lower lip lag of the upper lip) and for stutterer PC (specifically jaw and upper lip lag of lower lip) are much greater than the corresponding control subject latencies. For both of these subjects, post-therapy latencies are significantly reduced relative to their pre-therapy latencies, even though their post-therapy speech rate was significantly reduced compared to their pre-therapy rate. Turning next to sequential order, note that two of the stutterers, KH and PC, do not show the expected sequence in either the pre- or post-treatment condition. Also note that for stutterer KH, the pre- and post-treatment comparison shows a complete sequence reversal. Similar results were obtained for /pet, fit, and fet/ and indicate that post-therapy fluently enhanced speech can be marked by improved interarticulator relative-timing and, less frequently, by alteration of the sequence patterns, although the altered sequence may not be like that of the controls.

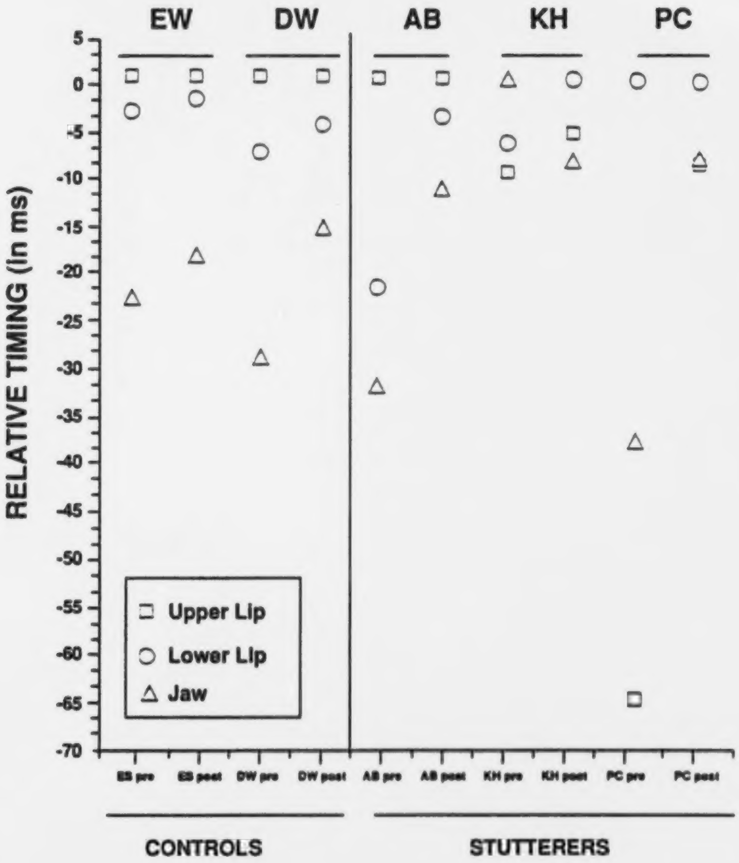


Figure 12. **Taken from Story and Alfonso (1989). Temporal organization of upper lip, lower lip, and jaw for /pit/.**

Turning next to the influences of phonetic context and speaking rate on the UL-LL-J sequence observed in /p/ closure, Figure 13, taken from De Nil and Abbs (1989b), shows the frequency of this sequence for repeated productions of /bae/. Speech rate varied from two to seven syllables per second over three second intervals. Note first that the UL-LL-J sequence for /b/ occurred in only 36 percent of the control subjects' utterances and in 25 percent of the stutterers' utterances, which is a much lower percentage than the /p/ closure sequence reported by Gracco and Abbs (1986) and Caruso et al. (1988) for control subjects. Figure 14 shows that, for both controls and stutterers, the expected sequence occurred most often at fast rates and least often at slower rates. Similar results were reported by Story, Alfonso, and Munhall (1987) for two normal talkers. The prevalent sequence, the UL-LL-J, occurred 84 and 89 percent of trials at normal and fast rates respectively for /p/ closure in /pit/ and /pet/ tokens. However, the prevalent sequence at slow rates was LL-UL-J, which occurred 69 percent of the time. Token-to-token variability for interarticulator relative-timing, particularly when involving jaw latency, was also much greater at slow rates compared to normal and fast. Similar rate dependant results were obtained for stutterers by Story (1990), even in the post-therapy condition when the stutterers successfully increased the frequency of their perceptually fluent speech.

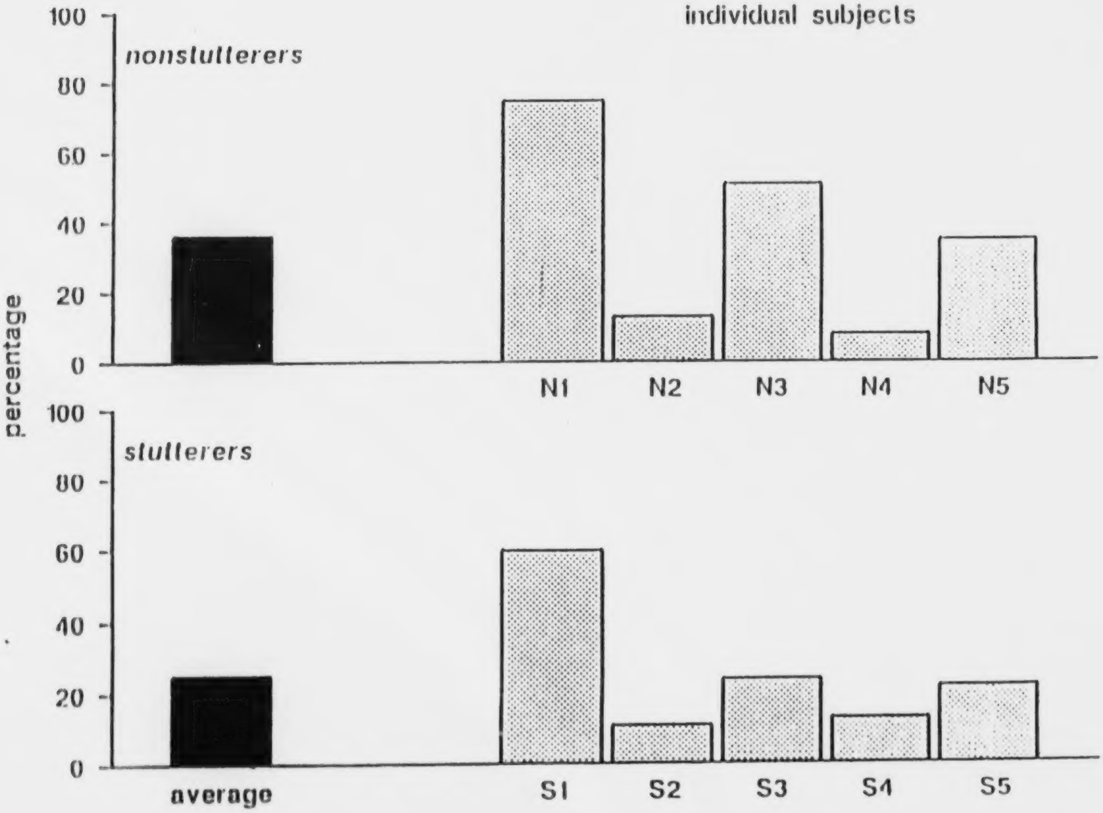


Figure 13. Taken from De Nil and Abbs (1989b). Frequency (%) of UL-LL-J sequence.

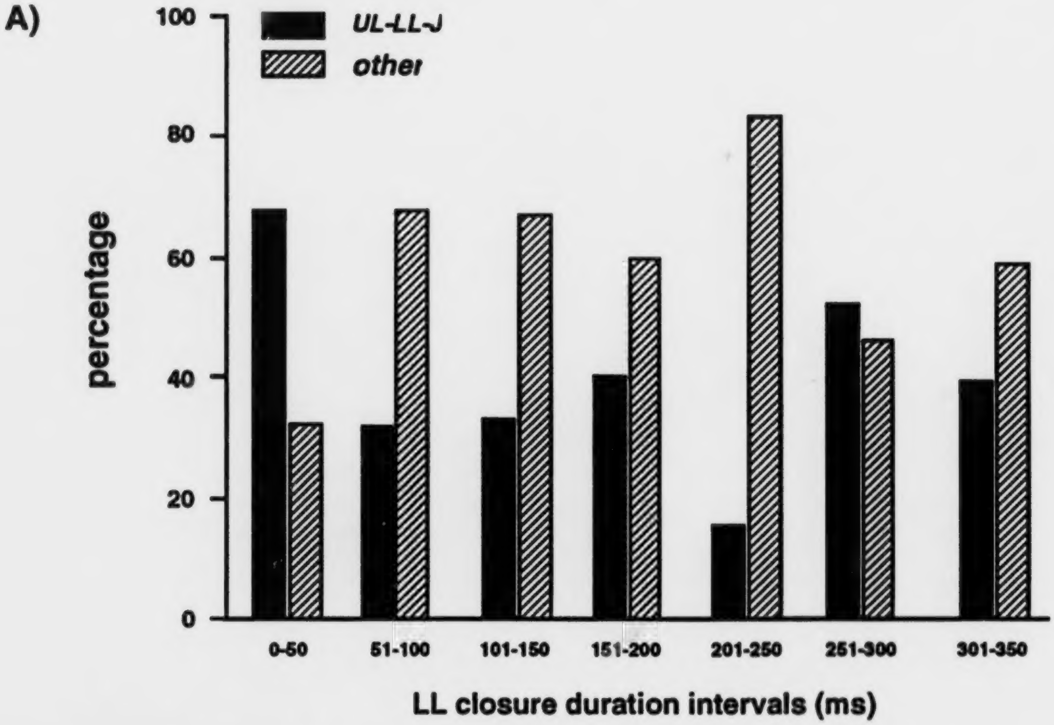


Figure 14a. **Taken from De Nil and Abbs (1989b). Nonstutterers: Peak velocity sequencing for each LL duration.**

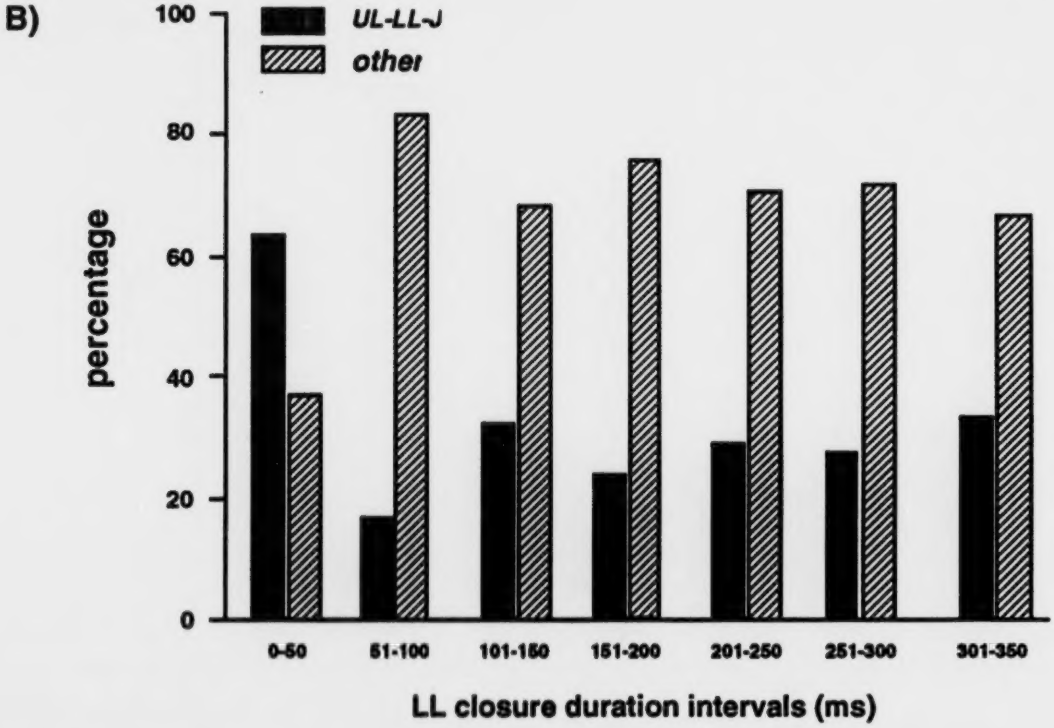


Figure 14b. **Taken from De Nil and Abbs (1989b). Stutterers: Peak velocity sequencing for each LL duration.**

CONCLUSION

This review suggests that stutterers' speech motor performance can significantly depend upon: (1) speech rate, (2) phonetic context, (3) stuttering severity, and (4) therapeutic history. These dependencies result in large variability in virtually all types of data collected from stutterers, and in certain cases, from control subjects as well. Obviously, we need to increase our data base, particularly with regard to speech motor control, before we can conclude with some degree of certainty whether velocity profiles, indices of motor equivalence, or just about any other measure of stutterers' perceptually fluent speech movements are normal, or alternatively, aberrant in some way. It would be to our advantage to begin a study of the feasibility of establishing a central store of physiological data with international access.

Because of the four dependencies listed above and the resultant variability problem, we should take care to adequately describe the essential methodological details of our experiments so that appropriate inferences can be made across different experiments, particularly because different protocols might reflect different dependencies. For example, it might be appropriate to limit the phonetic context, as in the example of the well studied initial /p/ in "sapapple" to address certain research questions, or it might be appropriate in a different experiment to compare one kinematic parameter across a variety of phonetic contexts. Similarly, it might be appropriate to gather data from limited recording sites from a limited number of subjects depending on the specific question the experiment seeks to resolve. We need to continue to weigh the advantages and disadvantages of these protocols with regard to the hypotheses under test, and to clearly describe all of the essential details related to these four dependencies.

In conclusion, the experiments reviewed here are particularly significant for the following two reasons. First, the results of these experiments, which examined speech motor organization during stutterers' perceptually fluent productions, indicate that the stutterers' speech motor system exhibits generalized spatio-temporal dysfunction rather than functioning normally except during moments of overt dysfluency. Second, the results of these experiments support the notion that the disorder of stuttering, as interpreted by McClean (1990) and Prins (1991), is associated with a relatively rudimentary form of motor control that is generally responsible for, and follows certain general principles of, spatio-temporal organization of all skilled motor activities, for example, handwriting and typing. This implies that the disorder of stuttering is associated with, in addition to the linguistic factors that are encoded in speech movements and the behavioral influences which might mediate them (e.g., Smith & Weber, 1988; Zimmerman, 1980c) irregularities at a rudimentary and pre-linguistic level of motor control.

Task-dynamic Modeling of Stutterers' Perceptually Fluent Utterances

The complexity of stutterers' speech motor control requires the synthesis of data stemming from new approaches and different disciplines. In this light, I conclude with a brief description of our first attempts to use the computational gestural model, and importantly, the dynamic component of the model, to test the assumptions about stutterers' perceptually fluent speech drawn from our tongue-jaw x-ray microbeam study (Alfonso et al, 1987a,b). Recall that we observed that the two stutterers produced alveolar gestures primarily by jaw displacement with relatively little tongue tip or tongue blade elevation. To model this behavior, the articulator weights associated with movements of the jaw were modified so that minimal anterior tongue displacement (with respect to the jaw) is produced during execution of alveolar closure and release gestures; virtually all of the net motion is achieved by jaw displacement. Under these constraints, multiple repetitions of normal rate alveolar stop-vowel-stop sequences are generated to determine if the acoustic output is perceived as normally fluent speech. Keep in mind that we have not modified other kinematic parameters. Rather, these utterances were generated with normal spatial and temporal coordination between the tongue and the jaw, only the relative magnitudes of the individual displacements have been modified. Figure 15 shows the results of these modifications and illustrates that the current version of the model allows us to begin this work. The top record of Figure 15 shows the acoustic signal associated with /etete/, one of the utterances used in the Alfonso et al. (1987a,b) experiment. The second record shows the tract variable movement, the boxes indicating the period during which the alveolar gesture is activated. The two lower records show the trajectories associated with jaw and independent tongue tip vertical displacement. The thick line in the jaw record represents the trajectory generated by the default articulator weights referred to as the normal condition, and shows that the model produces this utterance with jaw elevation that is appropriate for the vowel context and with only slight vertical assist for /t/ closure and release.

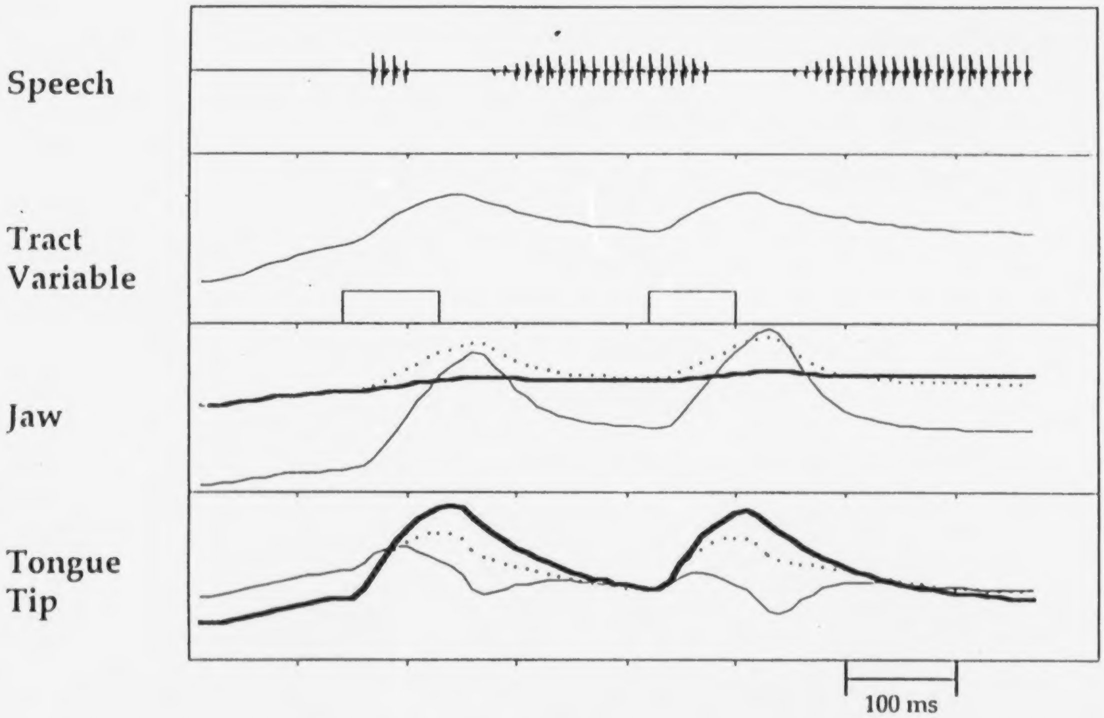


Figure 15. **Jaw and tongue tip trajectories for various modifications of the articulator weights associated with jaw movements. Thick lines represent trajectories generated in the normal condition. Dashed and thin lines represent increasing jaw contribution to alveolar closure and release, and the compensatory action of tongue tip displacement. See text for further details.**

The thick line in the tongue tip record shows that alveolar closure and release, in the normal condition, is achieved primarily by large displacements of the tongue tip. The dashed lines in the jaw and tongue tip records show the results of decreasing the articulator weights for the jaw and the resultant compensation of the tongue tip. Under this condition, the jaw component continues to assist the tongue body in vowel production but also clearly assists the tongue tip in /t/ closure and release. The acoustic output is indistinguishable from the normal condition. Finally, thin lines show the consequences of decreasing the articulator weights to obtain maximal vertical displacement of the jaw, and the compensatory adjustments to the tongue tip and body, until acoustic output is no longer possible. This condition corresponds most closely to the stutterers' tongue-jaw relative displacements observed in the Alfonso et al. (1987a,b) experiment. Note that the jaw contributes even further in alveolar closure and release, however, as mentioned above, this extreme jaw displacement does not result in an acoustic output, presumably because of the jaw's effect on other articulator movements, for example, the lips and tongue body. Our modeling of stutterers' speech is just beginning, but we are encouraged by our first attempts. We hope to be able to demonstrate the benefits of modeling five years from now at the third Nijmegen conference on speech motor control.

REFERENCES

Alfonso, P. J., Story, R. S., & Watson, B. C. (1987a). The organization of supralaryngeal articulation in stutterers' fluent speech production: A second report. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 21*, 117-129.

Alfonso, P. J., Story, R. S., & Watson, B. C. (1987b). Spatial-temporal variability of tongue-jaw movements in stutterers' fluent speech. *Asha, 29*-10, 159.

Alfonso, P. J., Watson, B. C., Baer, T. (1984). Muscle, movement, and acoustic measurements of stutterers' laryngeal reaction times. In M. Edwards (Ed.), *Proceedings of the 19th Congress of the International Association of Logopedics and Phoniatrics* (Vol. II, pp. 580-585). Perth, Scotland: Danscott Print Limited.

Alfonso, P. J., Watson, B. C., & Baer, T. (1987). Measuring stutterers' dynamical vocal tract characteristics by x-ray microbeam pellet tracking. In H. F. M. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (pp. 141-150). Wien: Springer-Verlag Press.

Alfonso, P. J., Watson, B. C., Seider, R. A. (1986). Tongue and jaw articulation in stutterers' fluent speech. *Asha, 28-10,* 166.

Baer, T., & Alfonso, P. J. (1984). On simultaneous neuromuscular, movement, and acoustic measures of speech articulation. In R. G. Daniloff (Ed.), *Articulation assessment and treatment issues.* San Diego: College Hill Press.

Baken, R. J., McManus, & Cavallo, S. A. (1983). Prephonatory chest wall posturing in stutterers. *Journal of Speech and Hearing Research, 26,* 444-450.

Bloodstein, O. (1987). *A handbook on stuttering*. Chicago, IL: The National Easter Seal Society.

Caruso, A. J., Abbs, J. H., Gracco, V. L. (1988). Kinematic analysis of multiple movement coordination during speech in stutterers. *Brain, 111,* 439-455.

Cooper, J. A. (1990). Research needs in stuttering: Roadblocks and future directions. *ASHA Reports, 18.* Rockville, MD: ASHA Publications.

De Nil, L. F. & Abbs, J. H. (1989a). Alterations in lower lip and jaw kinematics of stutterers and nonstutterers as a function of speech rate. *Journal of the Acoustical Society of America, 85-1,* S148. Paper presented at the Spring Meeting of the Acoustical Society of America. Syracuse, New York.

De Nil, L. F., & Abbs, J. H. (1989b). Influence of rate on stutterers' and nonstutterers speech movement timing. *ASHA, 31-10,* 65. Paper presented at the Annual Convention of the American Speech-Language-Hearing Convention. St. Louis, Missouri.

Gracco, V. (1988). Timing factors in the coordination of speech movements. *Journal of Neuroscience, 8,* 4628-4639.

Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research, 65,* 156-166.

Hutchinson, J. (1975). Aerodynamic patterns of stuttered speech. In L. M. Webster & L. C. First, (Eds.), *Vocal tract dynamics and dysfluency.* New York: Speech and Hearing Institute.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E., & Kay, B. (1986). A quantitative analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the acoustical Society of America, 77,* 266-280.

Lewis, J. J. (1975). *An aerodynamic study of 'artificial' fluency in stutterers.* Unpublished doctoral dissertation, Purdue University.

McClean, M. D. (1990). Neuromotor aspects of stuttering: Levels of impairment and disability. In J. Cooper (Ed.), Research needs in stuttering: Roadblocks and future directions, *ASHA Reports, 18.* Rockville, MD: ASHA Publications.

McClean, M. D., Kroll, R. E., & Loftus, N. S. (in press). Kinematic analysis of lip closure in stutterers' fluent speech. *Journal of Speech and Hearing Research.*

Peters, H. F. M., & Boves, L. (1988). Coordination of aerodynamic and phonatory processes in fluent speech utterances of stutterers. *Journal of Speech and Hearing Disorders, 31,* 352-361.

Prins, D. (1991). Theories of stuttering as event and disorder: Implications for speech production processes. In H. F. M. Peters, W. Hulstijn & C. W. Starkweather (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier Science Publishers.

Shapiro, A. I. (1980). An electromyographic analysis of the fluent and dysfluent utterances of several types of stutterers. *Journal of Fluency Disorders, 5,* 203-231.

Smith, A., & Weber, C. (1988). The need for an integrated perspective in stuttering. *ASHA, 30-10,* 32. Paper presented at the Annual Convention of the American Speech-Language-Heating Association.

Story, R. S. (1990). *Articulatory, laryngeal, and respiratory movements in stutterers' fluent speech before and after therapy.* Unpublished doctoral dissertation, University of Connecticut.

Story, R. S., & Alfonso, P. J. (1989). Temporal reorganization of lip and jaw gestures following an intensive therapy program. *ASHA, 31-10,* 65. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association. St. Louis, Missouri.

Story, R. S., Alfonso, P. J., & Munhall, K. (1987). Lip and Jaw kinematics at various speaking rates. *ASHA, 29-10,* 123. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association. New Orleans, Louisiana.

Van Riper, C. (1982). *The nature of stuttering* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Watson, B. C., & Alfonso, P. (1987). Physiological bases of acoustic LRT in nonstutterers, mild stutterers, and severe stutterers. *Journal of Speech and Hearing Research, 30,* 434-447.

Zimmerman, G. (1980a). Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *Journal of Speech and Hearing Research, 23,* 95-107.

Zimmerman, G. (1980b). Articulatory behaviors associated with stuttering: Cinefluorographic analysis. *Journal of Speech and Hearing Research, 23,* 108-121.

Zimmerman, G. (1980c). Stuttering: A disorder of movement. *Journal of Speech and Hearing Research, 23,* 122-136.

FOOTNOTES

[*] Appears in H. F. M. Peters, W. Hulstijn & C. W. Starkweather (Eds.), *Speech motor control and stuttering* (pp. 79-100). Amsterdam: Elsevier Science Publishers. (1991).

[†] Also University of Connecticut, Storrs. Now Chairman, Department of Speech and Hearing, University of Illinois, Champaign.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 111-122

The Effect of Speech Motor Training on Stutterers' Speech Physiology*

Peter J. Alfonso,† Joseph S. Kalinowski,†† and Robin S. Story†††

We report a number of our experiments that compare within-subject respiratory, laryngeal, and supralaryngeal kinematics of stutterers immediately before and after successful completion of two intensive therapy programs, one that emphasizes highly structured and physiologically oriented clinical targets and one that does not. Among the goals is to determine the post-therapy kinematics strategies that are requisite to the production of perceptually fluent speech, the notion being that the most important kinematic modifications leading to fluency will be shared by all successful stutterers in their post-therapy perceptually fluent speech even though therapy specific clinical instructions differ. The results demonstrate that post-therapy improvement in fluency co-occurs with spatial and temporal adjustments in each of the three systems, and that some of the same adjustments occur across both therapy programs.

This research seeks to resolve the following four questions. The first is a continuation of our previous work and asks whether certain kinematic profiles associated with stutterers' perceptually fluent speech are distinct from those of normal speakers. If they are different, the second question asks whether stutterers' aberrant kinematic profiles can be modified as a result of speech therapy and, if so, whether they become like those of normal speakers. The third question seeks to verify at the physiological level which clinical targets are incorporated in post-therapy fluent speech. The final question asks whether the post-therapy modifications we observe are related solely to program specific clinical instruction or whether they are requisite to fluent speech production. To determine the answers to these questions, within-subject pre- and post-therapy kinematic comparisons of the respiratory system (using Respitrace inductive plethysmography), laryngeal system (using photoglottography), and the supralaryngeal system (using optoelectric tracking to monitor the movements of the lips and jaw) are made of stutterers who have successfully completed a variety of therapy programs.

Reported here are the results of the first two experiments that examined speech related kinematics from eight stutterers immediately before and after completion of either one of two intensive fluency programs and four control subjects. Program 1, the Summer Residential Stuttering Clinic of Geneseo, New York, represents a Van Riperian type of program which primarily emphasizes speech rate control, and Program 2, the Communication Reconstruction Center's (CRC) of New York City version of the Precision Shaping Fluency Program (PFSP), represents a highly structured physiologically oriented program. The comparisons across different therapy programs are primarily motivated by our attempts to differentiate therapy induced kinematic modifications from kinematic modifications that are requisite to improved fluency. For example, kinematic modifications that occur in Program 2 subjects who show post-therapy increased fluency but not in equally successful Program 1 subjects could be

---

Preparation of this paper supported in part by NIH Grant DC-00121 and BRS Grant RR-05596 awarded to Haskins Laboratories. Final version of this paper was completed during the first author's tenure as Fulbright Research Scholar, The University of Nijmegen, The Netherlands.

considered requisite to the achievement of Program 2 clinical targets but not necessarily requisite to increased fluency. On the other hand, a kinematic modification that occurs in all successful subjects, including those who completed Program 1, which does not stress physiological targets, could be considered a physiological requisite to improved fluency. All stutterers who took part in these experiments were diagnosed as severe or moderate pre-therapy and as mild post-therapy. Kinematic measurements included traditional motor control indices, e.g., sequential ordering of articulator movements, and those that more directly address the achievement of the various program clinical targets. Two paradigms were used: a variable-foreperiod simple reaction-time task and a paradigm that assesses relatively natural speech, the production of the phrase "he see CVC again" where "C" represents various stops and fricatives. Only fluent utterances, defined perceptually and physiologically, are discussed here.

Turning first to our reaction-time (RT) experiments, the results of our previous reaction-time studies are presented as brief background. Watson and Alfonso (1983, 1987) showed that quantitatively different respiratory and laryngeal behaviors underline stuttering severity and foreperiod acoustic RT effects. Severe stutterers showed both delayed initiation and inappropriate organization of respiratory and laryngeal events leading to phonation at all foreperiods, while mild stutterers differed from severe stutterers primarily at short by not long foreperiods.

Kalinowski and Alfonso (1987) examined the relationship between stuttering severity and RT further by comparing within-subject pre- and post-therapy modifications to acoustic RT from 21 stutterers enrolled in either Program 1 or 2 and four control subjects. Briefly, subjects were first presented with a warning cue and then a phonate cue. The warning cue was the onset of a steady-state computer generated vowel /a/. The offset of the synthetic vowel served as a phonate cue, which signaled the subject to phonate the vowel /a/ as quickly as possible. Thirteen foreperiods, which are represented by the duration of the synthetic vowel, were used and varied from 100 to 3000 ms. Acoustic RT was determined as the time period from the offset of the phonate cue to the first glottal pulse in the acoustic record.

Figure 1 shows the results for the four control subjects who were tested before and after a four week interval that corresponds to the average duration of both therapy programs. No significant difference was found at any of the 13 foreperiods as a function of the session 1 and session 2 time interval. Figure 2 shows the same results for 13 stutterers enrolled in Program 1. Recall that this program emphasizes variable rate control and not the achievement of specific physiological targets. Note that for this group, acoustic RT values improve at all foreperiods as a function of therapy and that the average improvement, the difference between the grand means, is 54 ms. The pre- and post-therapy RT comparisons were significantly different at all 13 foreperiods. Figure 3 represents data for eight stutterers who completed Program 2, which emphasizes a fixed slowed speech rate and physiological targets, such as "gentle phonation" onset behaviors. This group showed little or no improvement at the relatively short foreperiods; however, at foreperiods greater than 700 ms, post-therapy RT improved relative to the pre-therapy value. The difference between the pre- and post-therapy grand mean is 17 ms which is statistically significant. These and other results of this experiment show that acoustic RT improvement covaries with therapy type and the magnitude of post-therapy fluency improvement, and suggest that acoustic RT represents a dynamic measure of phonatory function rather than a fixed and presumably neurologically-based delayed latency.

One of the purposes of the physiological RT experiments is to determine the physiological bases for the covariation between acoustic RT and improved fluency shown in Figures 2 and 3. Acoustic RT is measured as before, from the offset of the phonate cue to the first glottal pulse in the acoustic record. Kinematic RT measurements were made in reference to the onset of the warning cue. Respiratory RT is measured at the time of abdominal compression, and laryngeal RT at the time of the peak velocity for the adduction gesture. Procedures were similar to those in the acoustic RT experiments except that subjects were required to phonate the vowel /i/, rather than /a/, as quickly as possible.

Thus far, data from three controls and four stutterers who completed Program 1 have been analyzed (Kalinowski, in progress). The results indicate that prior to therapy, stutterers were slower than nonstutterers with respect to acoustic, respiratory, and laryngeal RT's at both relatively short (100-900) and long (1100-3000 ms) foreperiods. After therapy, stutterers generally demonstrated RT improvement at all levels of measurement and across short and long foreperiods. For some stutterers, certain post-therapy RT values at long foreperiods approached normal values.

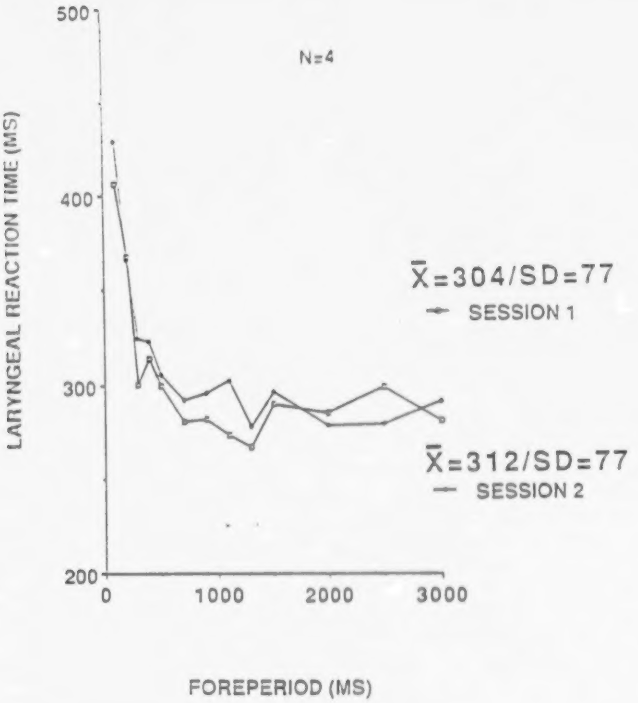


Figure 1. **Taken from Kalinowski and Alfonso (1987). Control subjects.**

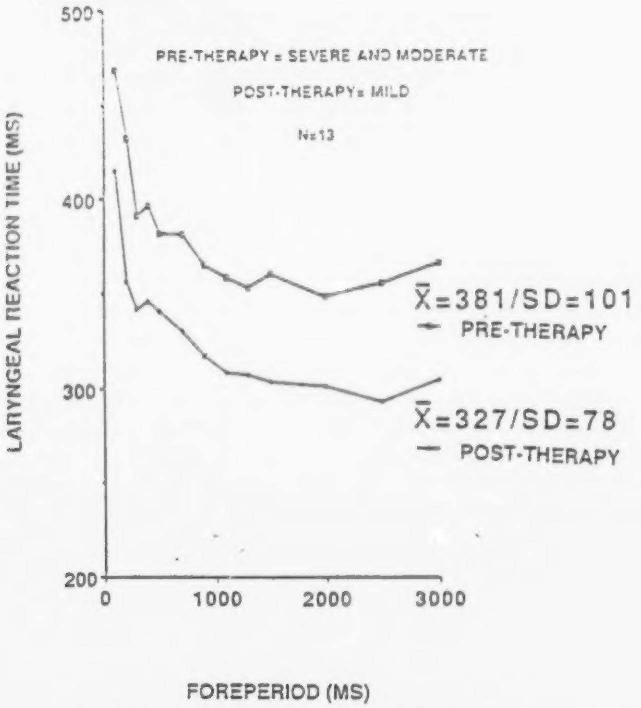


Figure 2. **Taken from Kalinowski and Alfonso (1987). Stutterers enrolled in Program 1.**

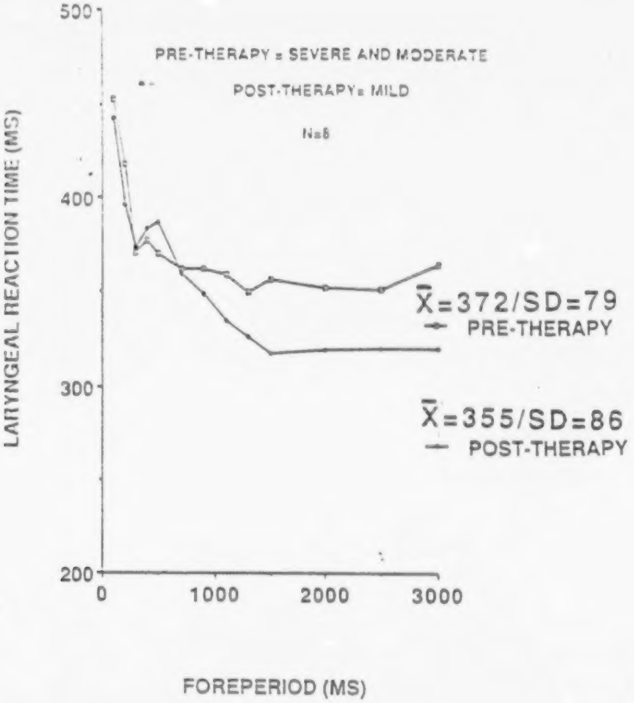


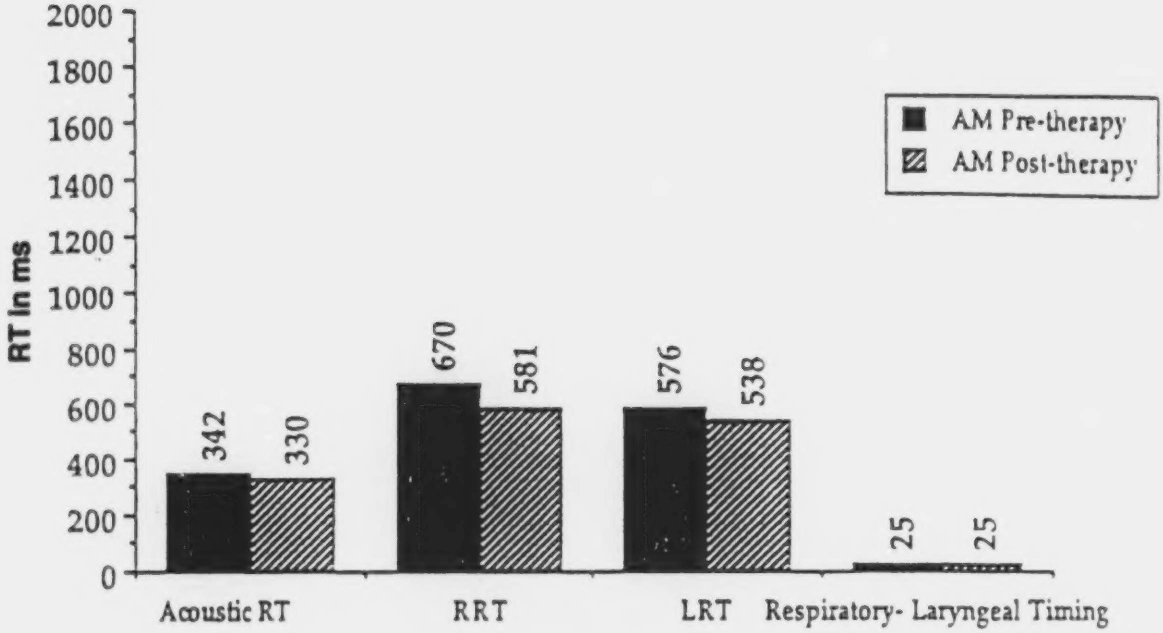
Figure 3. **Taken from Kalinowski and Alfonso (1987). Stutterers enrolled in Program 2.**

For example, Figure 4 shows pre- and post-therapy RT values at short (4a) and at long (4b) foreperiods for stutterer AM. The figure demonstrates that acoustic RT improved 12 ms at short foreperiods and 13 ms at long foreperiods. Note that both respiratory and laryngeal RT improved with acoustic RT, but unlike the acoustic RT profiles, the magnitude of respiratory and laryngeal RT improvement was much greater, approximately 200 ms at long foreperiods. Respiratory-laryngeal timing also improved for this subject, about 140 ms, but only at long foreperiods.

Figure 5 demonstrates: (1) even greater acoustic RT improvement for stutterer DLE (approximately 130 ms at short and 45 ms at long foreperiods) who correspondingly demonstrated greater fluency improvement than stutterer AM, and (2) that respiratory and laryngeal RT improvement was significantly greater at long foreperiods compared to short. However, this stutterer shows less improvement in respiratory-laryngeal timing compared to stutterer AM. For comparison, Figure 6 demonstrates comparable data for control TT, who does not show a significant RT-foreperiod interaction or a significant short versus long foreperiod effect for kinematic RT values.

The significant RT-foreperiod interaction for the stuttering subjects implies that the successful clinical strategies employed by Program 1 result in kinematic modifications, for example, increased respiratory charge and breathy phonation, that require a relatively long period of time to prepare and initiate. These and other data not shown here also demonstrate that respiratory-laryngeal temporal coordination improves at both short and long foreperiods for certain stutterers, although the magnitude of improvement is not always greater than the session 1 and session 2 variation observed for the control subjects we have analyzed thus far. This equivocal result implies that an understanding of the physiological bases for the covariation between acoustic RT improvement and perceptual fluency improvement is complex and requires an examination of other spatiotemporal kinematic modifications. For example, respiratory-laryngeal temporal coordination may make a greater contribution to improved acoustic RT than either respiratory or laryngeal RT in those stutterers who demonstrate either: (1) relatively short pre-therapy response latencies, and/or (2) appropriate levels of lung volume inflation for speech, and/or 3) appropriate laryngeal abductory/ adductory gestures for normal phonation.

4a)



4b)

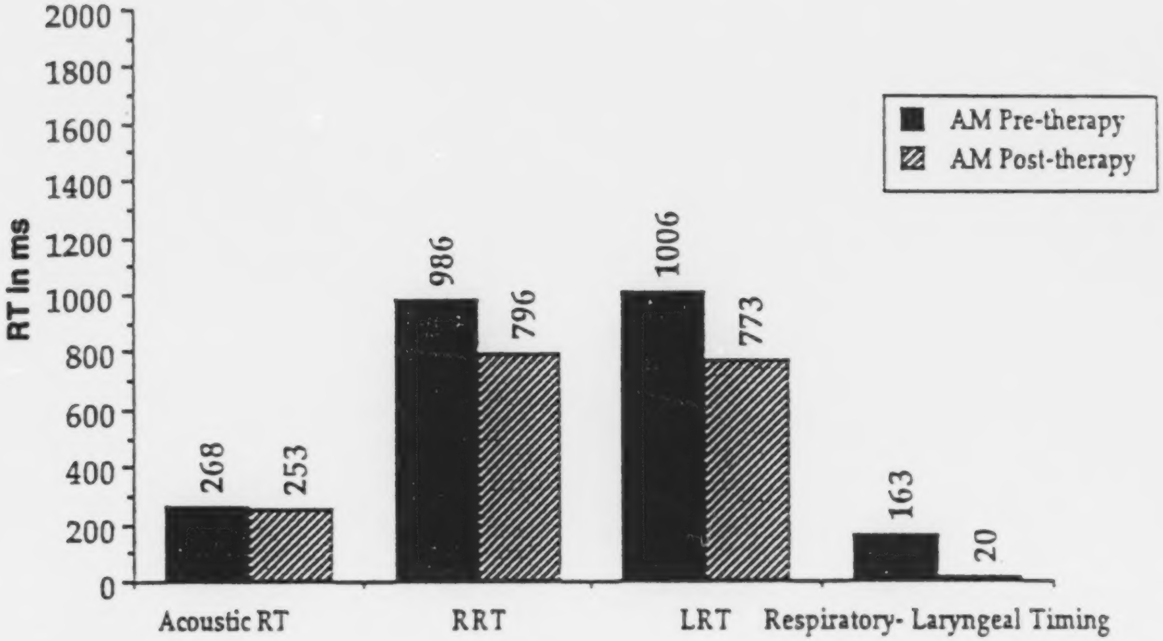
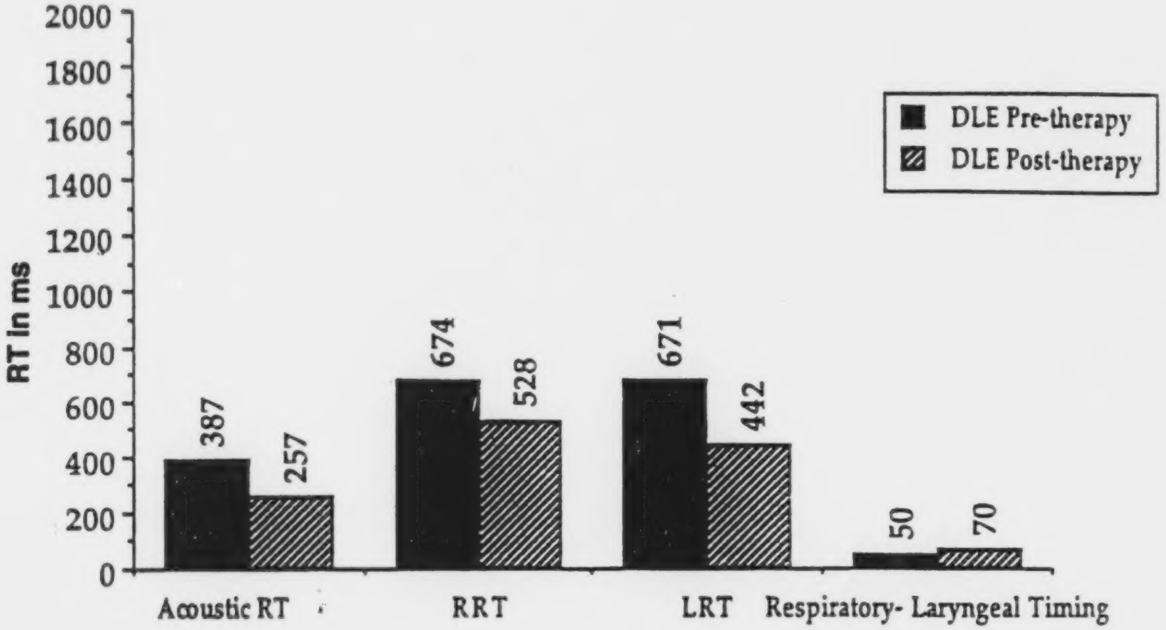


Figure 4(a). **Stutterer AM, pre- and post-therapy RT values at short foreperiods (100-900 ms).** ***(b).*** **Stutterer AM, pre- and post-therapy RT values at long foreperiods (1100-2000 ms).**

5a)



5b)

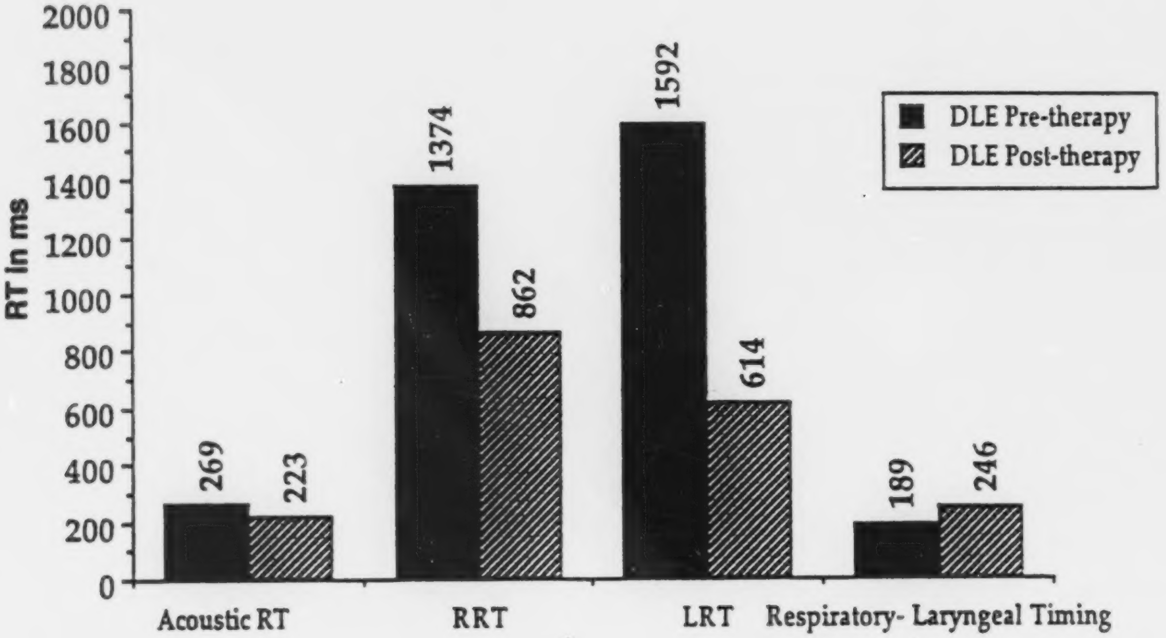
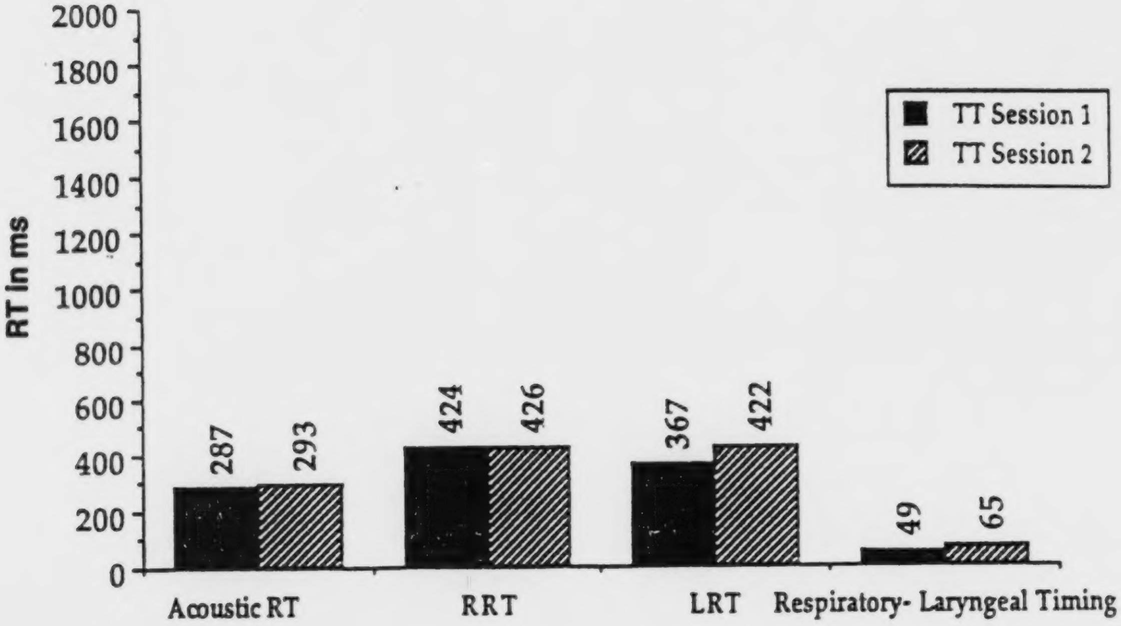


Figure 5(a). **Stutterer DLE, pre- and post-therapy RT values at short foreperiods (100-900 ms).** *5(b).* **Stutterer DLE, pre- and post-therapy RT values at long foreperiods (1100-2000 ms).**

6a)



6b)

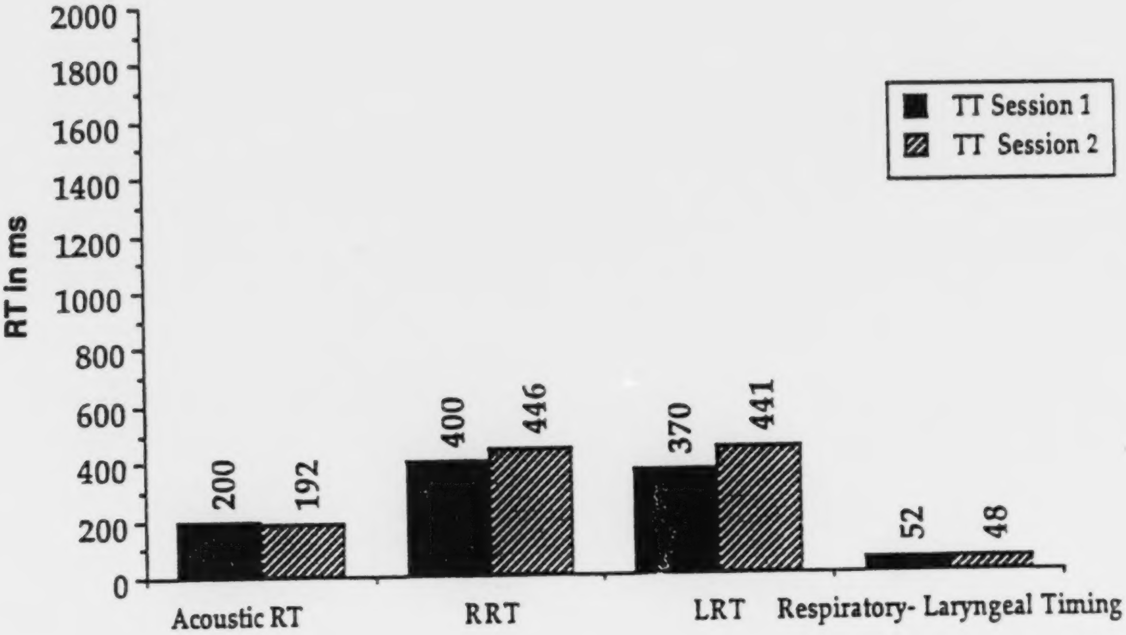


Figure 6(a). **Control TT session 1 and session 2 RT values at short foreperiods (100-900 ms).** *6(b).* **Control T, session 1 and session 2 RT values at long foreperiods (1100-2000 ms).**

We turn next to a discussion of the the second paradigm used in these experiments; that is, the kinematic modifications that occur in more natural speech conditions, such as in the carrier phrase "he see CVC again" where the target word is either "pete, pate, feet, or fate" (see Story & Alfonso, 1988; 1989; and Story, 1990). The data shown here are from stutterers who completed Program 2, which is the CRC version of PFSP. Subjects produced all utterances at three self-selected speech rates, normal, fast, and slow because the experiment sought to: (1) distinguish where possible between therapy induced post-therapy modifications with those modifications that are more natural precursors to perceptual fluency, and (2) determine the robustness of the post-therapy kinematic modifications and the saliency of the various clinical targets. Although the results indicate that post-therapy improved fluency co-occurs with kinematic modifications at all measured level of speech production at all speech rates, the following discussion is limited, due to time constraints, primarily to general kinematic modifications that occurred in the normal rate condition.

Figure 7 demonstrates expiratory volume, duration, and average flow for the two control subjects (shown on the left) and the three stutterers (shown on the right) at normal speech rates. Corresponding inspiratory data show similar patterns. All three stutterers significantly increased inspiratory and expiratory volume, duration, and flow. All of the stutterers presented similar post-therapy respiratory modifications at slow and fast rates. The two control subjects did not significantly increase inspiratory and expiratory volume and flow, from session one to session two.

Pre-speech respiratory abnormalities were observed prior to therapy, a typical pattern is shown in Figure 8. The summed respitrace signal and the PGG signal show that this perceptually fluent utterance was preceded by four inspiratory/expiratory cycles. These aberrant patterns were not identified audibly or visibly by the experimenters or reported by the stutterers at the time of the experiment. The frequency of similar pre-speech aberrant respiratory and phasic laryngeal behaviors decreased from 77 pre-therapy to 15 post-therapy across the three rate conditions. Reductions in respiratory abnormalities following therapy have also been reported by Murdoch, Killin, and McCaul (1989).

Respiratory measurements taken at pre- and post-therapy conditions were scaled with speech rate to determine whether the modifications in respiratory behaviors were speech rate motivated or were induced by other clinical targets. For example, the magnitude of the normalized pre- and post-therapy comparison for expiratory volume was greater on average for the stutterers, who increased volume by 38 percent, than it was for controls, who decreased volume by 25 percent. In a similar analysis, for stutterers the relative change in normalized respiratory volume across the three rates during the pre-therapy run was less than the relative change in the normalized pre- versus post-therapy comparisons at each rate. Post-therapy laryngeal modifications were also observed and suggest that stutterers' respiratory modifications were therapy induced. For example, post-therapy phonation was much more breathy and consequently would require greater expiratory volume. These results, taken with the significant reduction in aberrant respiratory behaviors, suggest that the changes in the control of the respiratory system that were observed in this experiment were not due entirely to speech rate changes across the treatment conditions, but rather were at least in part induced by other Program 2 clinical targets.

Turning next to modifications in laryngeal kinematics, Table 1 shows within-subject pre- and post-therapy comparisons for the duration of abduction for the initial consonant, the duration of adduction for the vowel, and the duration of the combined abduction-adduction laryngeal gesture for the four target words "pete, pate, feet, and fate." Note that the control subjects significantly decreased the duration of the laryngeal gestures. Stutterer KH showed a significant increase only in the case of the abduction gesture, while stutterers PC and AB showed significant increases in the duration of all gestures. Finally, note that there were no significant changes in the relative durations of the abductory or adductory changes for any of the subjects. For example, the table shows that for stutterer PC the duration of the abductory gesture was 45 percent of the total duration in the pre-therapy condition and 47 percent of the total duration in the post-therapy condition.

Certain comparisons were made to determine the relative contribution of speech rate and other clinical targets on the observed post-therapy laryngeal modifications. Considering the comparison with acoustic duration first, the two stutterers who showed significant increases in the laryngeal gestures shown here, stutterers PC and AB, did not show increases in the acoustic duration of the consonant. Rather, they decreased post-therapy speech rate by increasing vowel duration.

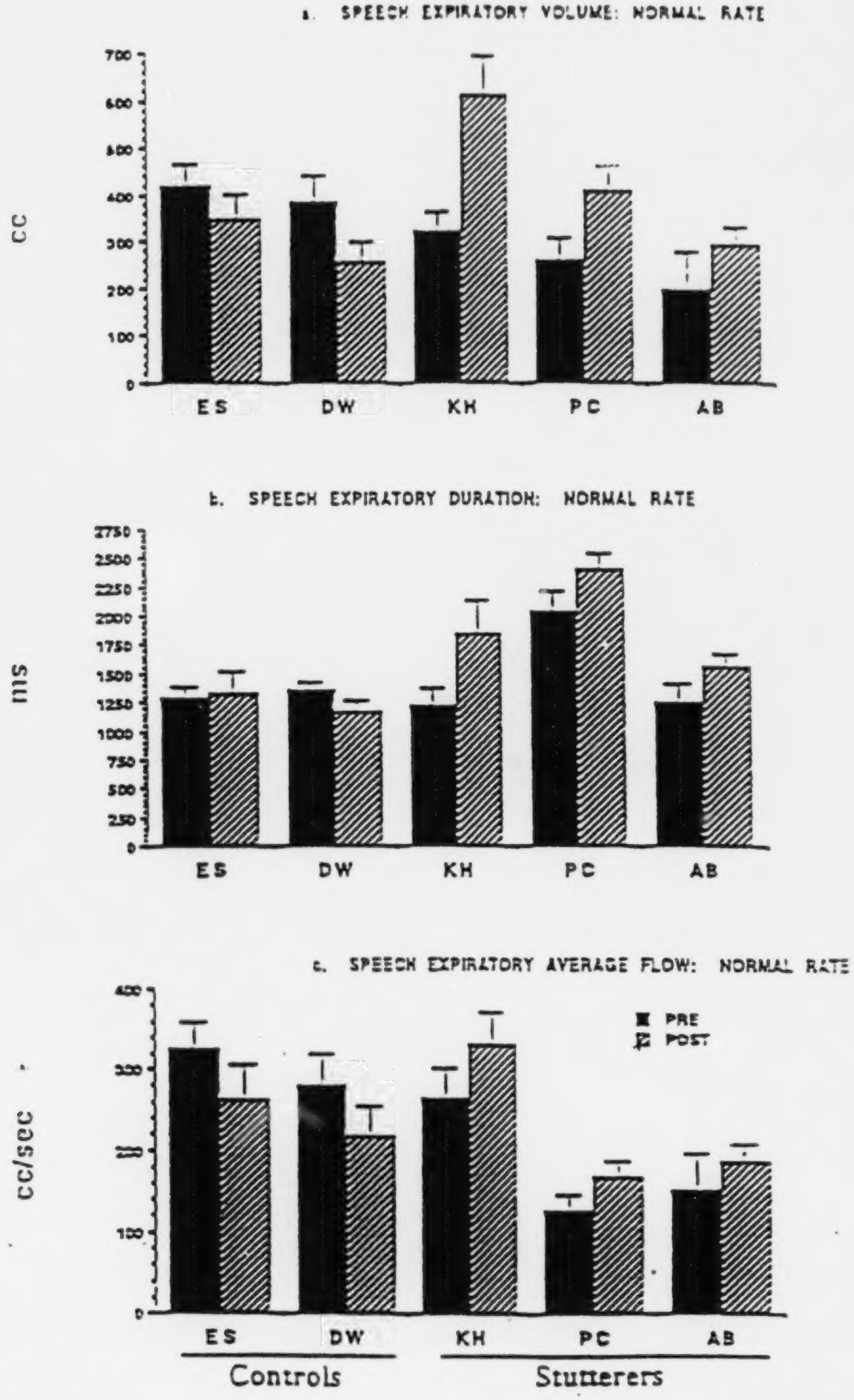


Figure 7. **Taken from Story (1990).**

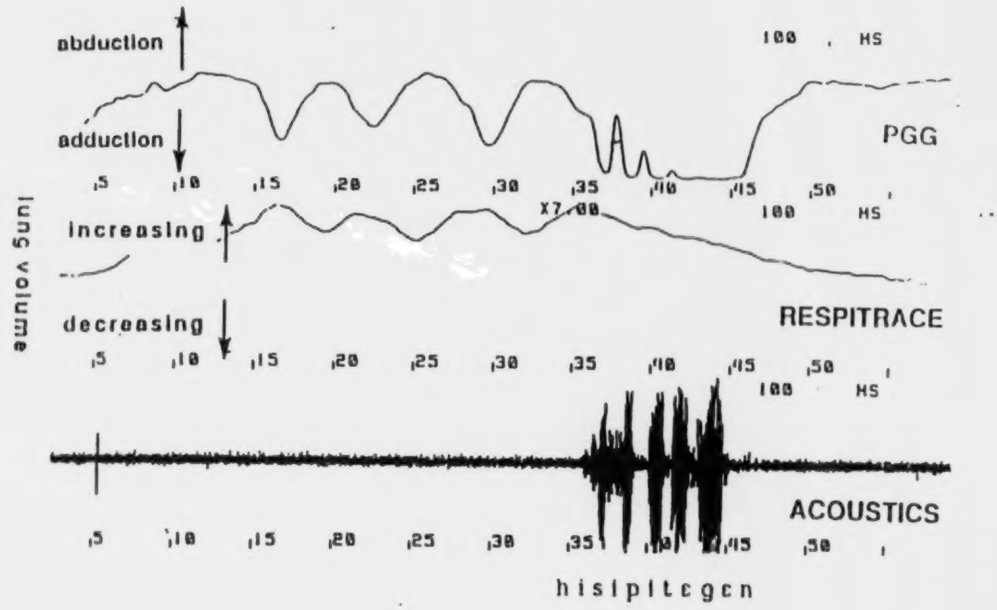


Figure 8. **Taken from Story (1990). Typical pre-therapy respiratory abnormality observed in stuttering subjects prior to perceptually fluent speech.**

Table 1. *Taken from Story (1990). Average duration (in ms) of abduction for the consonant, adduction for the vowel and total (abduction + adduction) laryngeal gesture for all words at the normal speech rate. The percent (%) total was derived from the abduction/total duration or the adduction/total duration.*

		ABDUCTION				ADDUCTION				TOTAL		
		n	mean	sd	% total	n	mean	sd	% total	n	mean	sd
CONTROL												
ES	pre	75	107.3	17.7**	.51	75	104.1	14.1**	.49	75	211.4	18.6**
	post	75	99.9	17.8	.52	75	91.6	10.7	.48	75	191.5	20.2
DW	pre	77	127.6	27.5**	.48	77	141.0	27.6**	.52	77	268.6	45.7
	post	71	107.9	16.0	.49	71	113.5	13.1	.51	71	221.4	23.7
STUTTERERS												
KH	pre	71	83.0	14.7*	.47	71	94.9	22.1	.53	71	177.9	18.9
	post	72	97.6	15.1	.52	72	91.3	23.1	.48	72	188.9	20.0
PC	pre	57	93.1	16.6**	.45	57	113.7	25.8**	.55	57	206.7	35.7**
	post	71	141.2	38.4	.47	71	158.1	26.0	.53	71	299.3	48.6
AB	pre	73	110.3	14.0**	.52	73	101.2	24.6**	.48	73	211.5	24.1**
	post	71	124.1	26.6	.50	71	124.7	25.1	.50	71	248.9	37.9

* = $p \leq .5$
** = $p \leq .01$

Similar comparisons were made with supralaryngeal data, for example, the duration of the lower lip gesture for the initial consonant. Although the magnitude and direction of change in the pre- and post-treatment comparison of the lip gesture is consistent with the magnitude and direction of change in the laryngeal gesture for the control subjects, the comparison does not always hold for the stutterers' data. For example, stutterer PC, who significantly increased the duration of all of the laryngeal gestures, decreased the duration of the lip gesture in three of the four target words. These and other data suggest that laryngeal modifications covary with improved fluency and occur at least in part as a function of clinical targets not associated with slowed speech rate.

Supralaryngeal measures included the displacement, peak velocity, and duration of the upper lip, lower lip, and jaw movements associated with closure and release for the initial consonant in the target CVC words. Significant post-therapy modifications were observed in the spatial and temporal domains for all stutterers. Some of these results are presented in a paper appearing elsewhere in this volume entitled "Implications of the concepts underlying task-dynamic modeling on kinematic studies of stuttering" and will be discussed only briefly here (Alfonso, 1991, this volume). For example, in the spatial domain, stutterers who demonstrated post-therapy fluency improvement reduced lip and jaw displacement (see Alfonso, 1991, this volume, Figure 8). In the temporal domain, stutterers who demonstrated fluency improvement post-therapy modified two temporal parameters of their post-therapy fluent speech. First, the most apparent temporal modification appears to be the improvement in inter-articulator relative-timing, in cases where the pre-therapy latencies were greater than expected. Most importantly, the reduction in interarticulator latencies occurred even though the stutterers' average phrase duration increased by about 225 ms post-therapy. Second, the patterns of the articulator movement sequences were modified post-therapy, in cases where the pre-therapy sequence was different than expected, although the modified sequence patterns were not similar to those demonstrated by control subjects (see Alfonso, 1991, Figure 14).

Various aspects of respiratory-laryngeal and laryngeal-supralaryngeal spatial and temporal coordination have been measured. Since an example of respiratory-laryngeal temporal coordination is given in the discussion of the reaction-time experiments, an example of laryngeal-supralaryngeal temporal coordination is given here. Table 2 shows latencies and movement sequences relative to the velocity peak associated with lower lip release for the initial consonant in the four target words with the velocity peak for the laryngeal abduction gesture for /p/ and /f/. As expected, all subjects showed a differentiation between stop and fricative production. That is, /p/ was produced with a lower lip release preceding the laryngeal abduction gesture, whereas /f/ was produced with the laryngeal abduction gesture preceding the lower lip release gesture. This sequence pattern has been observed in a relatively large group of normal speakers (Munhall & Löfqvist, 1987). In this figure, the frequency of lower lip lead or lag is shown by positive and negative numbers, respectively. For example, for control ES, all 18 productions of /pit/ were produced with a lip lowering gesture that preceded the laryngeal abduction gesture and all 18 productions of /fet/ were produced with a lip lowering gesture that lagged the laryngeal abduction gesture. A pre- and post-therapy comparison of the inter-system latencies and sequence movement patterns showed no differences between stutterers or controls for any of the target words.

In conclusion, these results suggest that post-therapy improvement in fluency co-occurs with spatial and temporal adjustments of the respiratory, laryngeal, and supralaryngeal systems. It is less clear whether modifications in the coordination of gestures between these systems occur, at least for the comparisons made here, particularly in regard to laryngeal-supralaryngeal temporal coordination. Some of the kinematic modifications we observe appear related to the clinical strategies associated with specific therapy programs, while others do not and may be related to normal motor control strategies, such as alterations in speech rate. Finally, some of the post-therapy kinematic modifications we observed are in the direction of normal speech kinematics observed in the control subjects.

Our plan is to compare the kinematic modifications of stutterers who successfully complete a variety of different therapy programs, the notion being that the most important modifications leading to fluency will be shared by all successful stutterers even though the clinical instructions to the different groups can differ. In this way, we hope to identify those kinematic strategies that are requisite to the production of perceptually fluent speech.

Table 2. *Taken from Story (1990). Mean latency values (in ms) between the time of peak laryngeal abduction and time of the lower lip release for the consonant /p/ or /f/ in the four stimulus words produced at normal speech rate. Positive values indicate that the lower lip release preceded the peak laryngeal abduction and negative values indicate that the peak laryngeal abduction occurred prior to the lower lip release.*

		/pIt			/pet/			/fIt/			/fet/		
		n	mean	sd	n	mean	sd	n	mean	sd	n	mean	sd
CONTROL													
ES	pre	18+	64.9	(10.4)	19+	57.4	(7.7)	17-,1+	-11.3	(10.8)	18-	-23.2	(11.6)
	post	16+	39.4	(7.0)	12+	30.6	(7.7)	17-	-27.4	(8.8)	17-	-24.6	(8.9)
DW	pre	17+,3	-10.3	(14.4)	19-	-14.4	(11.4)	19-	-39.1	(12.9)	17-	-43.5	(16.9)
	post	17+	29.2	(10.5)	6+,8-	-0.4	(10.9)	19-	-22.2	(9.0)	16-	-26.6	(9.7)
STUTTERERS													
KH	pre	18+	28.6	(10.3)	12+	40.2	(9.4)	17-	-37.9	(12.5)	17-	-53.3	(16.3)
	post	17+	35.5	(11.7)	18+	41.0	(10.0)	15-	-42.2	(21.1)	18-	-39.7	(15.3)
PC	pre	12+	41.0	(16.4)	10+	50.2	(31.3)	12-,4+	-33.4	(38.2)	10-,1+	-33.8	(22.9)
	post	14+	48.3	(11.9)	15+	41.1	(13.4)	19-	-97.6	(16.6)	17-	-78.9	(17.6)
AB	pre	15+	439.9	(17.0)	10+,5-	-5.5	(25.0)	14-,4+	-14.5	(18.4)	14-	-56.0	(15.7)
	post	15+	66.3	(14.0)	9+,8-	-3.8	(27.6)	16-	-67.9	(36.4)	18-	-86.4	(34.4)

REFERENCES

Alfonso, P. J. (1991). Implications of the concepts underlying task-dynamic modeling on kinematic studies of stuttering. In H. F. M. Peters, W. Hulstijn, & C. W. Starkweather (Eds.), *Speech motor dynamics and stuttering*. Wien/New York: Springer-Verlag.

Kalinowski, J. S. (in progress). *Kinematic strategies associated with enhanced fluency following completion of an intensive Van Riperian therapy program*. Unpublished doctoral dissertation, University of Connecticut.

Kalinowski, J. S., & Alfonso, P. J. (1987). Improvement in laryngeal reaction time with improvement in fluency. *ASHA, 29-10, 124*. Paper presented at the annual convention of the American Speech-Language-Hearing Association. New Orleans, LA.

Munhall, K., & Löfqvist, A. (1987). Gestural aggregation in speech. *PAW Review*. (A technical report of the perceiving-acting workshop on ecological psychology: The University of Connecticut), 2, 13-15.

Murdoch, B. E., Killin, H., & McCaul, A. (1989). A kinematic analysis of respiratory function in a group of stutterers pre- and post-treatment. *Journal of Fluency Disorders, 14*, 323-350.

Story, R. S. (1990). *Articulatory, laryngeal, and respiratory movements in stutterers' fluent speech before and after therapy*. Unpublished doctoral dissertation, University of Connecticut.

Watson, B. C., & Alfonso, P. J. (1983). Foreperiod and stuttering severity effects on acoustic laryngeal reaction time. *Journal of Fluency Disorders, 8*, 183-205.

Watson, B. C., & Alfonso, P. J. (1987). Physiological bases of acoustic LRT in nonstutterers, mild stutterers, and severe stutterers. *Journal of Speech and Hearing Research, 30*, 434-447.

FOOTNOTES

*Appears in H. F. M. Peters, W. Hulstijn & C. W. Starkweather (Eds.), *Speech motor control and stuttering* (pp. 513-526). Amsterdam: Elsevier Science Publishers. (1991).

†Also University of Connecticut, Storrs. Now Chairman, Department of Speech and Hearing, University of Illinois, Champaign.

††University of Connecticut, Storrs. Now at Dalhousie University, Halifax, Nova Scotia.

†††University of Connecticut, Storrs.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 123-140

Vowel Duration and Closure Duration in Voiced and Unvoiced Stops: There Are No Contrast Effects Here*

Carol A. Fowler†

The experiments test directly a claim in the literature [Kluender, Diehl & Wright, *Journal of Phonetics*, 1988] that the generally inverse relationship between durations of vowels and of closures for following voiced and voiceless obstruents occurs because language communities exploit an ostensible property of the auditory system identified as "auditory durational contrast." In this account, the long vowel before voiced as compared to voiceless consonants makes their already short closures sound even shorter than they are, thereby enhancing the perceptual distinctiveness of voiced and voiceless obstruents. The first experiment of the present series shows that contrast does not affect judgments either of closure durations or of vowel durations in VCV synthetic-speech disyllables. To the contrary, long vowels are associated with increased judgments that a closure is long, and, compatibly, long closures are associated with increased judgments that a vowel is long. A second experiment fails to replicate findings by Kluender et al. showing that classification judgments of nonspeech square-wave VCV analogs varying in the duration of the first (vowel-like) tone and in the duration of the (consonant-like) gap between tone pairs exhibit durational contrast. Judgments of these signals, like judgments of VCVs themselves, show assimilation, rather than contrast, effects. Whatever the reason may be why vowel durations and closure durations vary inversely in voiced and voiceless obstruents, it is not because language communities are exploiting durational contrast in the auditory system. For stimuli in these durational ranges and with these acoustic characteristics, there is no durational contrast effect to exploit. Moreover, if the assimilation effects that durational judgments did exhibit do, in fact, reflect distorting properties of the auditory system for signals such as VCVs, language communities are, evidently, snubbing those properties entirely in settling on durational properties of vowels before voiced and voiceless obstruents.

In a great many languages (but perhaps not all; see, e.g., Keating, 1985, and perhaps not in all contexts; see Davis & Summers, 1989), vowels are durationally longer before voiced than voiceless consonants, while consonant closures are shorter for voiced than for voiceless obstruents (e.g., Luce & Charles-Luce, 1985). In a recent paper, Kluender, Diehl, and Wright (1988) review and reject previous accounts of the so-called "vowel-length effect" and propose and test an account of their own.

Kluender et al. ascribe not only the vowel duration difference, but also its inverse relation to the following closure duration difference, to an exploitation by languages of a durational contrast effect in the auditory system. As they point out, contrast effects are found to be pervasive in perceptual judgments (so that, for example, a hefted weight is judged heavier if a lighter rather than a heavier weight has just been hefted (Guilford & Park, 1931); a tone is judged higher in pitch if it follows a lower than a higher pitched tone (Cathcart & Dawson, 1928-1929), etc. (See Warren, 1985, for a review.) A durational contrast effect would manifest itself for example as a decrease in the judged duration of an interval preceded by a longer than a shorter duration sound. Kluender et al. speculate that, in the context of a preceding long vowel, a consonant closure sounds shorter than in the context of a

---

This research was supported by NICHD Grant HD01994 to Haskins Laboratories.

preceding shorter vowel. Languages may exploit the auditory system's ostensible tendency to exhibit durational contrast in order to enhance the somewhat subtle differences between voiced-voiceless obstruent pairs. The long vowel before a voiced consonant makes its already short closure interval sound even shorter than it is; the short vowel before a voiceless consonant makes its already longer closure interval sound even longer.

Kluender et al. offer a test of their account. Two groups of listeners participated in their experiment, one group listening to /aba/-/apa/ disyllables, and the other listening to square-wave analogs of the disyllables. The purpose of the square-wave condition was to obtain an assessment of auditory-system judgments of nonspeech intervals. If judgments of speech disyllables and nonspeech stimuli pattern similarly, the comparison would lend support to a view that judgments of the speech intervals do not arise in a special-to-speech perceptual system (for research carried out using similar logic see, for example, Diehl & Walsh, 1989; Parker, Diehl, & Kluender, 1986; Pisoni, Carrell, & Gans, 1983). There were two continua in both stimulus sets that differed in the duration of the first vowel (or first tone in the square-wave set). The short and long vowel or tone continua had 155 and 245 ms initial vowels (tones) respectively. Members of each continuum differed in stop-closure (gap) duration, which ranged from 20 ms to 110 ms in ten steps. Short and long continua were blocked in the experiment and counterbalanced over subjects. In each half session, subjects first heard the endpoints of the appropriate continuum and were trained, using feedback, to hit one response button to one continuum endpoint and a second button to the other endpoint. Following an 80-item identification test of the endpoints, subjects took a 100-item test in which each continuum member occurred 10 times. They were to hit one or the other response button to classify each sound they heard.

Findings were that subjects hit the response button corresponding to the short-gap end of the continuum more frequently in the long-first-vowel (or -tone) condition than in the short-first-vowel (-tone) condition. Kluender et al. conclude as follows:

> We suggest that the principle of durational contrast provides a natural explanation both of the speech and nonspeech results. Specifically, a long initial segment makes a given medial gap seem shorter by contrast and hence, in the case of speech, more like a voiced segment. A short initial segment, on the other hand, makes a medial gap seem longer (i.e., in speech, more voiceless). Thus vowel-length differences are a means of enhancing the perceptual distinctiveness of the closure-duration cue for consonant voicing contrasts. (p. 161)

Their conclusions are not yet warranted, however. One reason why the conclusions are not warranted concerns the "principle" of durational contrast. In their references to contrast effects, Kluender et al. do not cite a single investigation of auditory durational contrast. In my search of the literature, I uncovered a few investigations that do report contrast (e.g., Behar & Bevan, 1961; Walker, Irion, & Gordon, 1981), but they used stimuli longer than the vowels and consonants of spoken languages. Moreover, the one discussion I found of effects of the duration of a filled interval (such as the vowel or tone in the stimuli used by Kluender et al.) on a following open interval (such as the closure or gap) reported a positive not a negative relation between filled interval duration and judged open interval duration (Woodrow, 1928). However, this study, like the others I found, used stimuli somewhat longer than the vowels and consonants in the experiments by Kluender et al. Finally, there are grounds provided by the literature on contrast effects for guessing that there should be no contrastive effect of a vowel on a following closure duration.

Kluender et al. cited Warren's (1985) review and theoretical paper on contrast effects (which Warren calls "criterion shifts") to make the point that such effects are extremely pervasive in perceptual judgments. However, they did not mention that Warren's interpretation of these effects does not predict a criterion shift in closure duration judgments due to variation in vowel duration. Warren proposed that criterion shifts are consequences of the way that perceivers use past experience with relevant objects or events of some type to make judgments about present tokens; that is, a criterion shift is a consequence of the way that perceptual learning works. He suggested that perceivers make perceptual judgments by assessing a characteristic of an object or event (for example, the weight of an object) relative to their past experiences with other relevant objects of the same type, and, rationally, they weight very recent experience disproportionately (see Bjork, 1974 and Freyd & Johnson, 1987, for similar ideas about "memory averaging"). In an experiment in which stimuli vary along some dimension, subjects behave as if they are maintaining a running average of the values along the varying dimension, but recent trials contribute disproportionately to the average. In an experiment on weight judgments, a recent encounter with a very heavy weight will temporarily strongly affect the assessment of the typical weight of objects in the experiment (Guilford & Park, 1931); accordingly a lighter weight encountered next is judged, relatively, very light indeed. Analogously, in judgments of vowel duration, a very long vowel judged recently should strongly affect a judge's running assessment of typical vowel duration, and a following shorter vowel will be judged even shorter than if it had not been preceded by the long vowel. This kind of effect should give rise to trial-to-trial contrast effects in judgments of speech stimuli (as it does, for example, in judgments of vowel color; see e.g., Repp, Healy, & Crowder, 1978). However, it should not give rise to effects of vowel duration on following closure duration. Warren reviews evidence suggesting that criterion shifts are restricted to influences of past experience with objects or events of the same type. This is consistent with the idea that these effects are not blind sensory distortions of perceptual information, but rather reflect adaptive uses of past experience in perceptual judgment. In the example Warren uses, Johnson (1944) found contrast effects of past judged weights on current judgments; however, if, during the experiment, subjects happened to lift an object, such as a book, that was not one of the experimental stimuli, that object had no effect on subsequent weight judgments. Analogously, successive vowel durations may contribute to a running assessment of typical vowel durations in some experiment, and consonant durations may contribute to a running assessment of typical consonant durations, but a long vowel should not change the running assessment of consonant duration, at least not vial the criterion shift proposed by Warren (1985).

All of this would be moot, of course, had the experiments of Kluender et al. convincingly demonstrated a durational contrast effect for VCVs and squarewave stimuli. However, their test of the effect was indirect in failing to ask subjects to judge closures (gap durations) explicitly. While it is true that, within a stimulus set, closure or gap duration was the only acoustic dimension to undergo variation, that in itself does not ensure that subjects made closure- or gap-duration classifications. Subjects were not told how the stimuli differed in a test; rather, they had to determine that for themselves during the feedback trials with endpoint stimuli. Naive listeners may not even be aware that stops include a silent or near-silent closure interval; if they were unaware of the closures in the speech stimuli, they would have responded on some other basis in the speech condition. Indeed, possibly they made /b/-/p/ classifications, with the short-gap signals sounding like /aba/ and the long-gap signals sounding like /apa/. If so, then the speech condition itself simply replicated previous findings of a vowel-length effect on voicing classifications (e.g., Raphael, 1972).[1]

In respect to the speech/nonspeech comparison, Kluender et al. argued that they demonstrated, in addition, that the basis for the vowel-length/voicing relationship is to be found in the auditory system of the listener. Elsewhere (Fowler, 1990) I have attempted to challenge the logic by which such inferences are drawn based on speech/nonspeech comparisons of the sort made by Kluender et al. I will make an analogous attempt here in the General Discussion. However, the first experiment focuses specifically on the vowel-length/closure-duration relation in speech and asks whether there is a durational contrast effect there.

EXPERIMENT 1

In the present experiment, three groups of listeners made classification judgments of a set of VCV disyllables. Two of the three groups made discrimination judgments as well. The VCV disyllables differed in closure duration, vowel duration and post-release aspiration/voicing. One group was instructed to classify the closure durations as long or short (and to discriminate pairs based on closure duration); a second group was instructed to classify vowels as long or short (and to discriminate pairs based on vowel duration); the third group classified the consonants as "b" or "p."

Judgments of closure duration provide a direct test of the view that a durational contrast effect occurs in VCV disyllables. Judgments of vowel duration were included as a matter of interest to determine whether a contrast effect also might work in the opposite direction. Results from the three groups will perhaps permit inferences as to the bases for the button-press classifications made by subjects in the experiment of Kluender et al.

Methods

Subjects. Subjects were 36 students at Dartmouth College who participated for course credit. All were native speakers of English who reported normal hearing. Twelve students participated in each of the three groups in the experiment.

Stimulus materials. The VCVs were not designed originally to address the claims of

Kluender et al. Accordingly, they were not designed after the stimuli used in that research. They were, however, designed to be consistent with measured vowel and closure durations characteristic of vowel-/b/ or -/p/ sequences reported in the literature (e.g., Luce & Charles-Luce, 1985). Moreover, the closure-duration range falls almost entirely within that tested by Kluender et al., and the vowel duration range falls in between the short and long vowel durations they used.

Eighty 'VCəs were used in the identification tests of Experiment 1. The disyllables reflected the orthogonal combination of four initial-vowel durations, four closure intervals and five combinations of post-release aspiration/voicing. The 'VCəs were synthesized on the serial resonance speech synthesizer at Haskins Laboratories. Vowel durations ranged from 160 to 190 ms in 10 ms steps; closures ranged from 90 to 120 ms in 10 ms steps. At the /b/ end of the aspiration/voicing continuum, the AH parameter (noise excitation for formants) was set to zero during formant transitions at consonant release, while the AV parameter (glottal pulsing excitation for formants) was set to 20 dB, rising to a value of 30 dB by the end of the transitions. At the opposite end of the continuum, the AH parameter was set to 20 dB, while the AV parameter was set to 0 during the release transitions. Three intermediate stimuli were created by varying the AH and AV parameters between their endpoint values in equal steps.

Steady-state formants for the first vowel (bandwidths 50, 100 and 200 Hz) were 730, 1090 and 2440 Hz. Transitions into the closure were 40 ms in duration. F1 fell to 400 Hz, F2 to 760 Hz and F3 to 2110 Hz. Rising transitions out of the closure were also 40 ms in duration; starting frequencies were the ending values of the falling transitions, and steady states for the final schwa vowel were 640, 1190 and 2390 Hz. The second vowel had a duration of 100 ms. Amplitude rose over the first 40 ms of the signal and remained steady until the closure, where it was set to zero; during release, the amplitude rose from 20 to 30 dB; it fell to zero over the final 40 ms of the vowel. Fundamental frequency was steady at 120 Hz during the steady-state of the first vowel; it fell during closing transitions to 104 Hz. During release and much of the steady-state of the second vowel, fundamental frequency was set at 120 Hz; it fell over the last 40 ms of the vowel to 100 Hz.

A single identification test was designed for presentation to all three experimental groups. The test presented each of the 80 disyllables (four vowel durations × four closures × five values of the voicing/aspiration variable) three times each in random order with the constraint that each item occur once in each third of the test. There were 3.5 seconds between trials and seven seconds between blocks of 24 trials.

It was impractical to include all 80 disyllables in the discrimination test. For that test, the aspiration/voicing continuum values were fixed at the /b/ end of the continuum, leaving 16 disyllables in which vowel and closure durations were varied orthogonally. In the discrimination test, all possible pairings of the 16 disyllables were included, yielding a 256-item test. There were 750 ms between disyllables of a pair, 3.5 seconds between trials and 7 seconds between blocks of 32 trials.

Procedure. Subjects were run in groups of 1-3 in a quiet room; they listened over headphones to stimuli presented at a comfortable listening level. They were assigned to experimental conditions on a rotating basis, depending on order of arrival. All subjects took the identification test first. Subjects in the closure- and vowel-duration groups took a following discrimination test.

Closure duration group. Subjects in this group were shown a pair of spectrographic displays of VCV disyllables differing in closure duration. One disyllable had the largest closure duration and one the smallest of the values used in the experiment. The disyllables had identical values of the vowel duration variable (the second to shortest value) and of the aspiration/voicing variable (the middle value). I explained to subjects the relation between the visual display and the disyllables they would be hearing; in particular, the formant bands were identified with the vowel and the empty gap with the consonant. (I explained that the interval was silent for /b/ and /p/ because the speaker's lips were closed.) The reason for showing the displays was to convince subjects that there was a silent gap in the syllables that might be judged. Listeners were told that they would be classifying the duration of the silent interval between the two syllables of each disyllable as long or short. They listened to a demonstration sequence consisting of the stimuli displayed in the spectrograms played in alternation five times each. Listeners were told that the first disyllable of each pair had the long closure and the second the short closure. They were instructed to listen to the alternating pairs attempting to attend to the relevant difference so that they could classify disyllables by closure duration in the test proper. They were allowed to hear the sequence additional times if they wished, and subjects in this group frequently did request a replaying. In the identification test, listeners were told that they would hear one disyllable on each of 240 trials and they were to classify the closure as long or short by writing L or S on the answer sheet, guessing if necessary.

After the identification test, listeners participated in the 256-item discrimination test. They were instructed to write "1" on their answer sheet if the first disyllable of each pair had the longer closure and "2" if the second disyllable had the longer closure. They were instructed to guess if necessary, not leaving any answers blank.

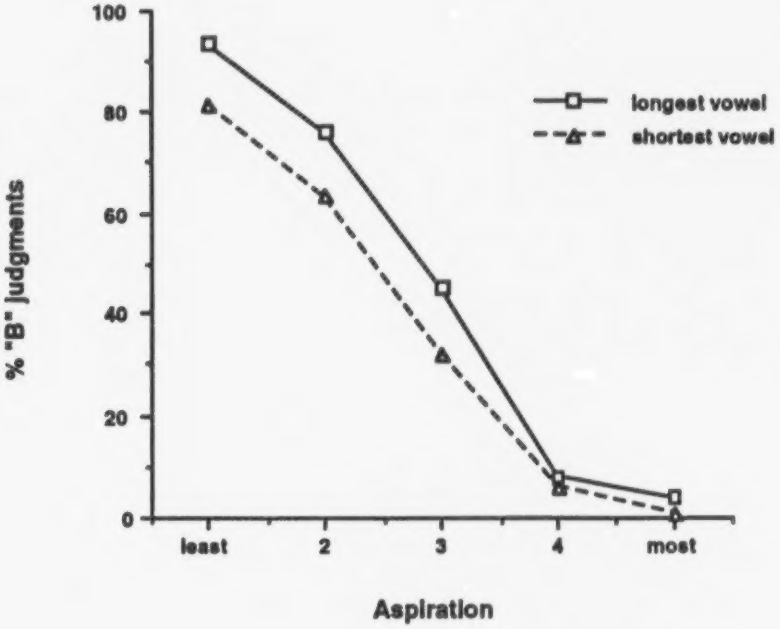
Vowel duration group. Subjects in this group were shown a pair of spectrographic displays of VCV disyllables differing in vowel duration. One disyllable had the longest vowel duration and one the shortest of the values used in the experiment. The disyllables had identical values of the closure duration variable (the second to shortest value) and of the aspiration/voicing variable (the middle value). The relation between the display and the disyllables they would be hearing was explained to this group as it was explained to the closure duration group. Listeners were told that they would be classifying the initial-vowel duration of the VCV disyllables as long or short. They listened to a demonstration sequence consisting of the stimuli displayed in the spectrograms played in alternation five times each. Listeners were told that the first disyllable of each pair had the long vowel and the second the short vowel. They were instructed to listen to the alternating pairs attempting to attend to the relevant difference so that they could classify disyllables by vowel duration in the test proper. They were allowed to hear the sequence additional times if they wished; however, in this group, such requests were rare. In the identification test, listeners were told that they would hear one disyllable on each of 240 trials and they were to classify the initial vowel as long or short by writing L or S on the answer sheet, guessing if necessary.

Listeners then went on to the discrimination test, writing "1" if the first disyllable of each pair had a longer initial vowel than the second disyllable and "2" if the second disyllable's initial vowel was longer. Again, listeners were instructed to guess if necessary, not leaving any blank responses.

"b"/"p" group. Listeners in this group participated only in an identification test. Their task was to classify the medial consonant as "b" or "p" guessing if necessary. Subjects in this group first heard alternating b/p disyllables having endpoint values on the voicing/aspiration continuum and identical values (next to shortest) on the other continua. Listeners heard the demonstration pairs five times each before proceeding to the identification test. They were told that the first disyllable of each pair was /aba/ and the second was /apa/. They were to listen to the pairs attempting to learn to label the consonants as "b" and "p," respectively. They were permitted to hear the demonstration sequence again if they wished; however, requests to do so were uncommon. On the identification test, using the same test sequence as the listeners in the other two groups heard, subjects' task was to identify the consonant of each disyllable as "b" or "p," guessing if necessary.

Results and Discussion

The aspiration/voicing continuum is not of particular relevance here; accordingly for analyses following the present one, identification data are collapsed over that continuum. However, pre-collapsed data are presented in Figure 1 from the condition in which subjects identified the consonants as "b" or "p" to illustrate that stimulus variables contributed in the expected way to voicing identification. Figure 1 presents a subset of the data from that condition. In Figure 1 (top) percent "b" responses are compared across the aspiration continuum for stimuli having the longest of the four vowel durations (averaged over closure duration differences) and for stimuli having the shortest of the four vowel durations. In Figure 1 (bottom), analogous data compare the effect of closure duration across the aspiration continuum. As expected consonants with little or no aspiration, with long vowels and with short closures are reported as "b"; those with opposite values are reported as "p." In an analysis of variance on all of the data from the "b"-"p" identification condition with factors Aspiration, Vowel duration and Closure duration; all main effects were highly significant (Aspiration: $F(4,44) = 134.45$, $p < .001$; Vowel duration: $F(3,33) = 17.16$, $p < .001$; Closure duration: $F(3,33) = 6.63$, $p = .0013$). "B" judgments decrease significantly with decreases in aspiration, with decreases in vowel duration and with increases in closure duration. Just one interaction reached significance, that between Aspiration and Vowel duration ($F(12,132) = 4.07$, $p < .001$). The interaction was significant because the effect of vowel duration on "b" judgments was diminished for aspiration continuum members 4 and 5 where "b" judgments approached the floor even with long vowels.



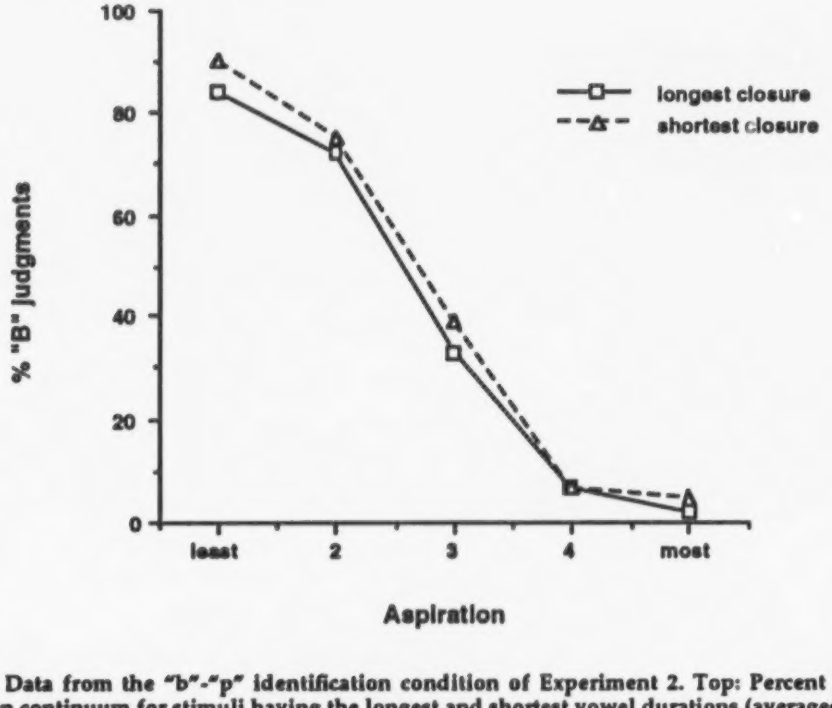


***Figure 1.* Data from the "b"-"p" identification condition of Experiment 2. Top: Percent "b" responses across the aspiration continuum for stimuli having the longest and shortest vowel durations (averaged over variation in closure duration). Bottom: Analogous plot for stimuli having the longest and shortest closure durations.**

Closure duration judgments. I next consider the identification judgments of the group of major interest in the experiment—listeners who classified closure intervals as long or short. If there is a durational contrast effect of the sort that Kluender et al. propose, then closures should be judged shorter preceded by longer vowels than shorter vowels.

Figure 2 shows the two significant effects obtained in an analysis of variance on average "long" classifications of closure durations. The figure plots the percentage of "long" closure judgments for stimuli differing in closure duration, but averaged over vowel duration (dark bars) and for stimuli differing in vowel duration averaged over closure durations (striped bars). The figure shows that "long" closure judgments increased both as closure duration increased *and* as vowel duration increased. In an analysis of variance with repeated measures factors closure duration and vowel duration, both main effects are significant while the interaction does not approach significance (closure: $F(3,33) = 11.63$, $p < .001$; vowel: $F(3,33) = 36.05$, $p < .001$). The vowel factor, in fact, accounts for considerably more of the variance in judged closure durations than does the closure variable itself (31.3% versus 16.6%). Tests for linear decreases in "long" judgments are also highly significant on both variables (closure: $F(1,33) = 34.39$, $p < .001$; $F(1,33) = 194.44$, $p < .001$). The effect of vowel duration on judged closure duration is the reverse of a durational contrast effect.

The results of the discrimination test reveal a similar effect. Table 1 presents proportions of trials on which the first presentation of the disyllable was judged to have the longer closure in the different trial types of the experiment. Vowel duration differences (duration of the first vowel in the disyllable pair [V1] minus duration of the second disyllable's first vowel [V2]) are displayed horizontally in the table while closure duration differences (Cl1-Cl2) are displayed vertically. Generally, percentages of judgments that the first disyllable has the longer closure increase as the closure duration difference increasingly favors the first disyllable (that is, top to bottom in the table). The difference in the percentage of "first longer" judgments between extreme pairs (Cl1-Cl2 = 30 ms versus Cl1-Cl2 = -30 ms) averages 37% overall in the table and 36% on disyllabic pairs having the same vowel duration (the middle column of Table 1).

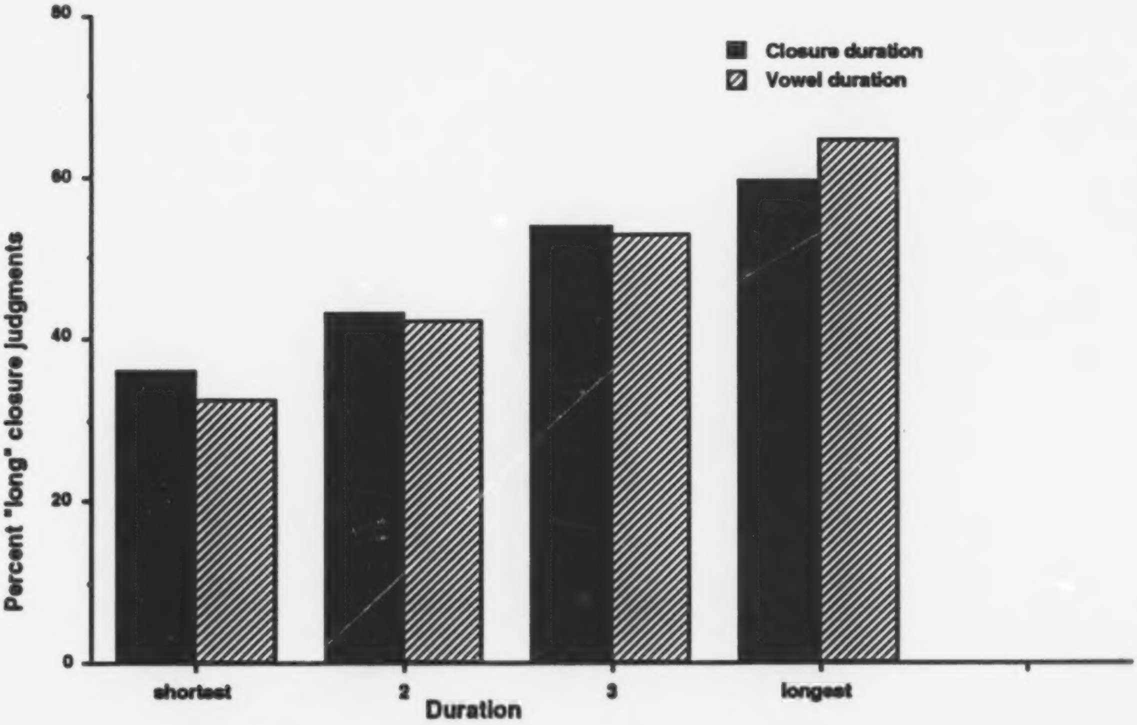


Figure 2. Classifications of closure durations as long or short in Experiment 1 as a function of variations either in closure duration itself or in vowel duration.

Table 1. *Percent judgments that the first disyllable had the longer closure in the discrimination test of Experiment 1. Across the rows is the vowel duration difference (V1-V2) in ms; down each column is the closure duration difference (Cl1-Cl2) in ms.*

		V1 - V2 (ms)						
		-30	-20	-10	0	10	20	30
	-30	08	04	28	29	42	54	**58**
	-20	08	19	26	43	63	**58**	83
	-10	11	17	39	48	**61**	72	69
Cl1-Cl2	0	21	33	40	**51**	67	76	83
(ms)	10	19	35	**51**	60	68	85	69
	20	29	**40**	61	65	72	81	83
	30	**42**	42	53	65	86	92	100

Looking left to right in the table reveals the effect of vowel duration on closure-duration discrimination judgments. There is a large effect, analogous to that in the identification test. If the first disyllable has a longer vowel than the second, subjects are biased toward judgments that the first *closure* is longer than the second. That the effect of the vowel duration difference on closure duration discriminations is, in fact, larger than the effect of the closure duration difference itself can be seen by looking at the diagonal printed in boldface in the table. Along the diagonal, the vowel duration difference and the closure duration difference between the disyllables of a pair cancel out, leaving disyllables in each pair that are identical in duration. There are, in fact, more judgments that the first closure is longer than the second when that difference is 30 ms in the wrong direction, but the vowel duration difference offsets it (top right in the table) than when the closure duration difference is 30 ms in the right direction, again offset by an opposite direction of difference in vowel durations. Likewise when there is no closure duration difference at all (the middle row of the table), judgments that the first closure is longer than the second increase from 21% to 83% as the vowel duration difference increasingly favors the first vowel. This compares to a shift from 29% to 65% in the middle column of the table where vowel durations are the same in disyllables of a pair.[2] In an analysis of variance on judgments that the "first disyllable's closure is longer" with factors closure duration difference (Cl1-Cl2) and vowel duration difference, both main effects are significant ($F(6,66) = 15.87$, $p < .001$; $F(6,66) = 55.79$, $p < .001$, respectively) with the vowel duration difference again accounting for more of the variance (37.2%) than the closure difference (9.9%). The interaction was nonsignificant.

Clearly in these stimuli there is no contrast effect of vowel durations on judgments of following closures. It is unlikely that subjects in the speech condition of the experiment of Kluender et al. were judging gap duration.

Vowel duration judgments. Comparable analyses were done on the vowel-duration classification and discrimination judgments. Relevant findings are presented in Figure 3 and Table 2.

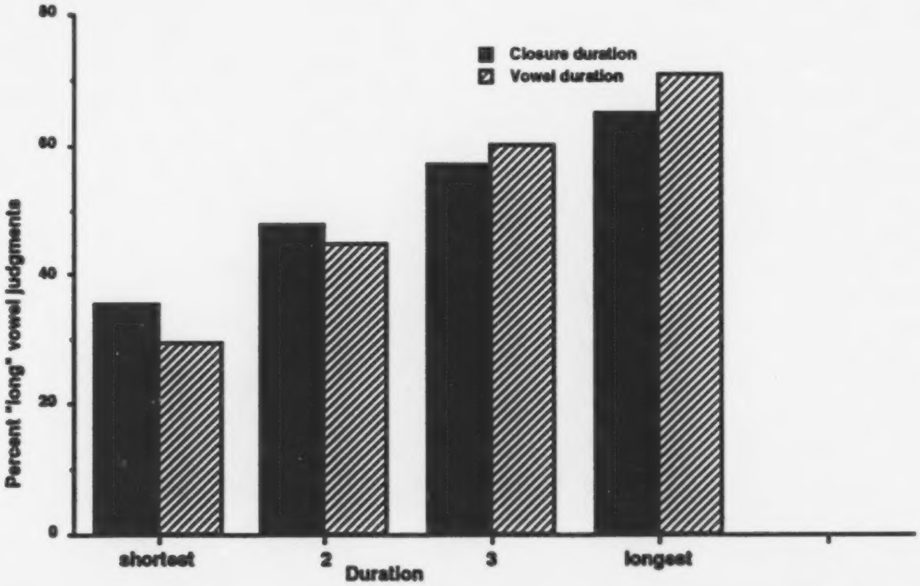


Figure 3. Classifications of vowel durations as long or short in Experiment 1 as a function of variations either in vowel duration itself or in closure duration.

Table 2. *Percent judgments that the first disyllable had the longer vowel in the discrimination test of Experiment 1. Across the rows is the vowel duration difference (V1-V2) in ms; down each column is the closure duration difference (Cl1-Cl2) in ms.*

		V1-V2 (ms)						
		-30	-20	-10	0	10	20	30
	-30	08	21	33	35	47	63	83
	-20	21	29	26	42	68	67	71
	-10	14	29	41	50	57	72	81
Cl1-Cl2	0	15	35	39	60	65	69	92
(ms)	10	47	29	46	56	77	89	81
	20	50	31	54	67	85	94	88
	30	25	46	61	65	94	83	92

Figure 3 shows the effects of closure duration and vowel duration on judgments of vowel duration. As in the previous analysis of classification judgments, both main effects are significant, while the interaction is not (closure duration: $F(3,33) = 24.49$, $p < .001$; vowel duration: $F(3,33) = 47.31$, $p < .001$) with the vowel effect stronger (19.6% and 41.3% of the variance accounted for respectively). Tests for linear decreases in judgments that the vowel was long were significant for both the closure-duration variable ($F(1,33) = 72.5$, $p < .001$) and the vowel-duration variable ($F(1,33) = 141.30$, $p < .001$). In short, vowels were more likely to be judged long if they were long or if they were followed by a long closure than if they were short or were followed by a short closure. The positive effect of long closures on judgments of the vowel durations as long was clearly audible in this test.

Once again the discrimination test bears out the findings on the identification test. Both the relative closure durations and the relative vowel durations of each disyllable pair affect the percentage of judgments that the first disyllable of a pair has the longer first vowel. In an analysis of variance the effects of relative closure duration ($F(6,66) = 11.81$, $p < .001$; 6.8% of variance accounted for) and of relative vowel duration ($F(6.66) = 56.17$, $p < .001$; 37.7%) were significant; the interaction was nonsignificant.

Analogously to the closure duration findings, there is an influence on judged vowel duration of the duration of a following closure, but it is a positive effect, not a contrast effect. Accordingly, it is unlikely that listeners in the speech condition of the experiment of Kluender et al. were judging relative vowel durations even though the vowels would be perceptibly different in duration due to differences in duration of the following closures.

"b"-"p" judgments. The purpose of obtaining "b"-"p" judgments was largely to ascertain that, in the disyllables of Experiment 1, long vowels and short closures were associated with increased "b" judgments. As reported above, an analysis of variance with factors aspiration, closure duration and vowel duration, all main effects were significant, while just one interaction, between aspiration and vowel duration was significant. Figure 4 shows the effect of vowel duration and closure duration on "b" judgments. Each variable shows the predicted effect, each with one deviation from monotonicity across variation in duration. Despite these individual departures from monotonicity, tests for linear increases in "b" judgments as closure duration decreased and as vowel increased yielded highly significant results ($F(1,33) = 10.03$, $p = .0034$; $F(1,33) = 37.44$, $p < .001$, respectively).

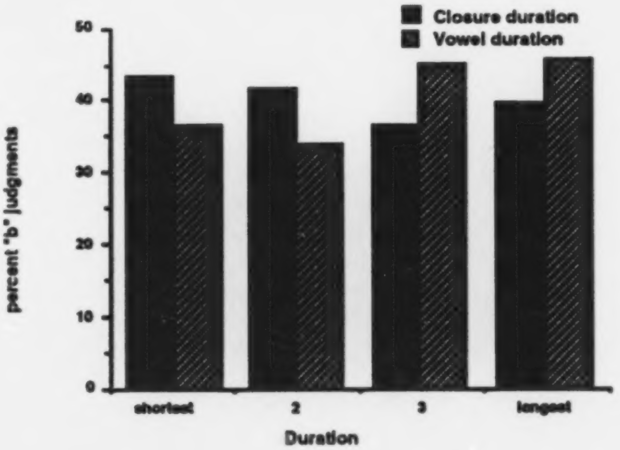


Figure 4. Identifications of consonants as "b" in Experiment 1 as a function of variations either in vowel duration or in closure duration.

This, then, is the only test in which the dependent measure (b/p judgments) showed the patterning of the findings of Kluender et al. As closure duration increased, "b" responses fell; as vowel duration increased, "b" responses increased (see also Figure 1). Perhaps, rather than responding to closure duration, the subjects of Kluender et al. made b/p classifications in the speech test.

This leads of course to the question of what perceived variable governed button pressing to the nonspeech stimuli in the research of Kluender et al. where subjects were unlikely to be making b/p classifications, but where their classifications of short- and long-gap stimuli patterned as the speech stimuli did. That issue is addressed in Experiment 2.

EXPERIMENT 2

The most obvious basis on which subjects in the nonspeech condition of Kluender et al. might make their button press responses is, of course, the acoustic difference in gap duration. Within a test, that is the only acoustic variable to undergo change. However, if gap duration is the basis for the button press classifications, then listeners apparently do not respond in the same way to variation in the duration of a preceding vowel and tone in their closure- and gap-duration classifications. Whereas Experiment 1 demonstrated a positive relation between vowel duration and judged closure duration, subjects in the nonspeech condition of Kluender et al. demonstrated a negative relation between initial tone duration and judged gap duration.

In the present experiment, I used square-wave stimuli designed after those of Kluender et al. and attempted to determine the conditions under which such stimuli give rise to a durational contrast effect.

Methods

Subjects. Subjects were 56 students at Dartmouth College, who received course credit for their participation. They reported normal hearing. Subjects were run until there were 18 students in each condition of the experiment. Data from two subjects were eliminated from one of the experimental conditions for reasons given below.

Stimulus materials. Stimuli were designed after the description provided by Kluender et al.[3] They were square-wave analogs of the stimuli used by Kluender et al. in their speech condition. Accordingly, two sets of 10 signals were created, each signal consisting of two tones separated by a silent gap. The sets were identical except that the first tone in one set was 155 ms long while that in the other was 245 ms. Amplitudes were invariant over the tones except for 10 ms rise and decay intervals. The fundamental frequency was set at 256 Hz except for a fall to 175 Hz during the last 40 ms of the first tone and a symmetrical rise from 175 to 256 Hz over the first 40 ms of the second tone. Within a set of 10 stimuli, the gap duration varied from 20 ms to 110 ms in 10 ms steps. Figure 5 displays a sample nonsense signal. (The figure does not resemble the schematized spectrograms in Figure 1 of Kluender et al. in certain respects. However, the spectrograms depicted in Kluender et al. also disagree with the description of the signals in the text. Keith Kluender (personal communication, January 17, 1990) has told me that the description in the text is accurate.)

The stimuli were digitized (filtered at 10 kHz and sampled at 20 kHz) and stored in a computer for presentation to listeners. Three test orders were devised for each stimulus set. A demonstration sequence presented the two continuum endpoints in alternation five times. Next, an 80-item test presented each endpoint 40 times in random order. Finally, a 100-item test presented all 10 continuum members ten times each with the constraint that each member occur once in each successive tenth of the test. There were 3.5 seconds between trials in the 80-item and 100-item tests.

Procedure. Subjects listened over headphones in a quiet room. Three groups of subjects were run. Each group heard both sets of test orders. Half of the subjects in each group took the tests involving the short-tone stimuli followed by the long-tone tests; the remaining subjects took the tests in the reverse order. Each test sequence took about 20 minutes; subjects took a five-minute break between the half-sessions.

A/B group. This group provided a pencil-and-paper approximations to the button press condition of Kluender et al. That is, subjects in this group were not told how the endpoint stimuli differed. They were told that each nonsense sound they would hear consisted of two tones separated by a gap. Initially, they would be hearing just two such sounds that differed audibly in a way that they would have to figure out for themselves. They were to call the first sound of the alternating pair in the demonstration sequence "A" (in fact, the short-gap endpoint) and the second sound "B." Subjects were invited to listen to the demonstration sequence as many times as they wished until they were comfortable that they could label the endpoints accurately.

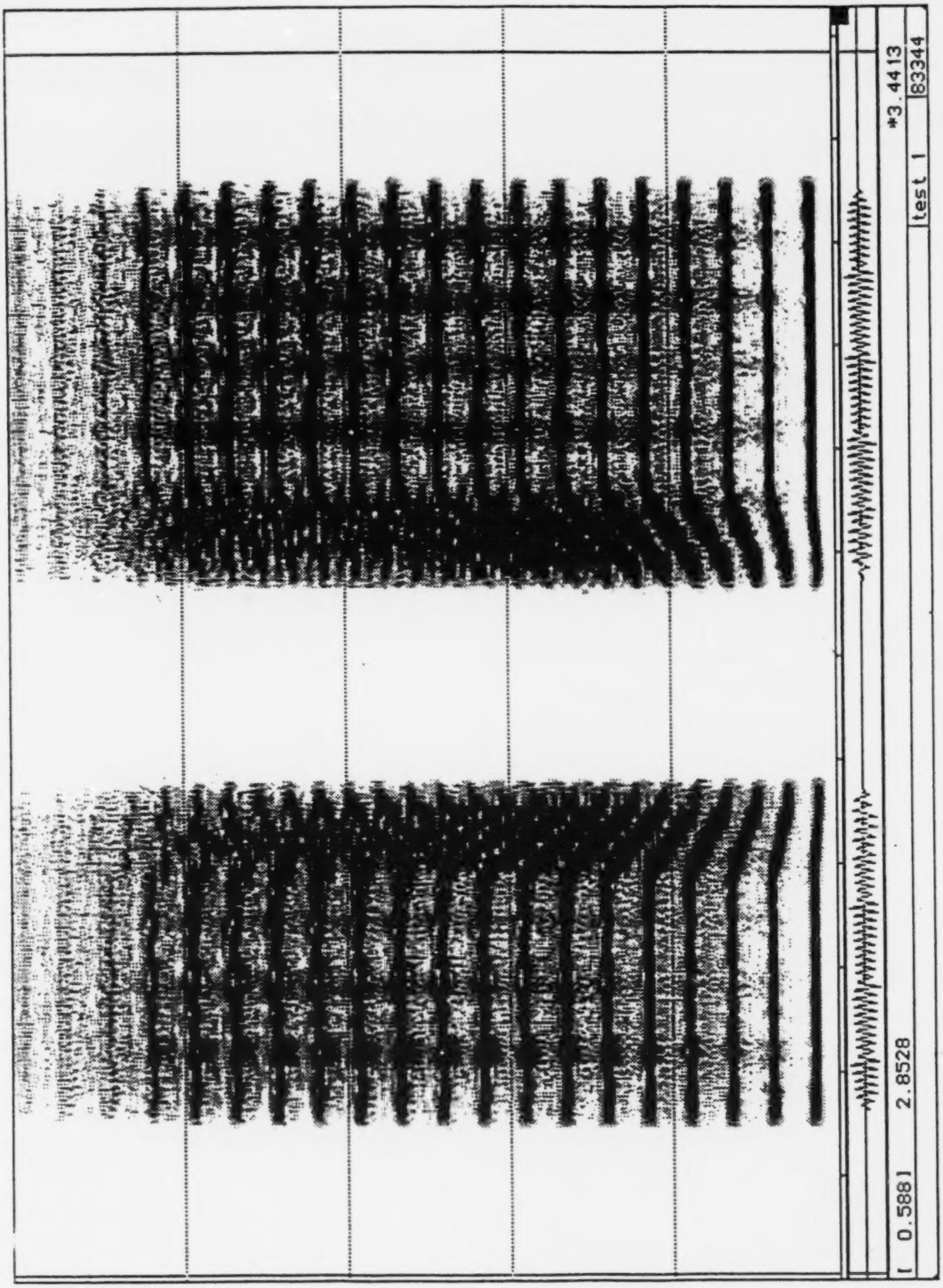


Figure 5. **A sample square wave stimulus.**

They went on to take the 80-item test on the endpoints. They were told to label each sound either A or B, guessing if necessary. (This procedure contrasts with that of Kluender et al. not only in that Kluender et al. obtained button-press responses, rather than having subjects write their responses, but, more importantly, they gave their subjects feedback after each trial. Our findings below suggest that neither procedural difference was important.) On the 100-item test, I told them that now they would be hearing not only the sounds they had been labeling but some new sounds as well. Their task was to write A if the sound they heard was more like A than like B and to write B if the sound was more like B; they were instructed to guess if necessary. After a five minute break, subjects went on to take the same series of tests on the other stimulus set.

L/S group. This group provided an explicit test of the durational contrast hypothesis. Subjects in this group were told that the nonsense sounds in the demonstration sequence consisted of two tones separated by a silent gap and that the difference between the sounds was in the duration of the gap between the tones of a pair. They were told that the first nonsense sound in the alternating sequence had a considerably shorter gap than the second. Their task was to learn to label the first nonsense sound "S" and the second one "L." Subjects were invited to listen to the demonstration sequence as many times as they wished until they were comfortable that they could label the endpoints accurately. They went on to take the 80-item test on the endpoints. They were told to label each sound either S or L, guessing if necessary. On the 100-item test, I told them that now they would be hearing not only the sounds they had been labeling but some new sounds as well. Their task was to write S if the sound they heard was more like the short-gap sound than like the long-gap sound and to write L if the sound was more like the long-gap sound; they were instructed to guess if necessary. After a five minute break, subjects went on to take the same series of tests on the other stimulus set.

B/P group. Kluender et al. report that the "square-wave stimuli, which replicate neither the harmonic nor the formant structure of speech, were judged by all three experimenters to bear little perceptual resemblance to the corresponding /aba/-/apa/ stimuli." (p. 158) Although I concur with this assessment, I ran a group anyway in which the task was to label the nonsense sounds as /aba/ (by writing "B") or as /apa/ ("P"). The reason for running such a group was that the corresponding group in Experiment 1 provided the only data conforming in pattern to both the speech and nonspeech data of Kluender et al. I presumed that the A/B group above would replicate the nonspeech findings of Kluender et al., but that the L/S group might not (in view of the finding in Experiment 1 that listeners judging closure duration did not). There would, then, remain the question as to what perceived variable provided the basis for the A/B (and button-press) classifications. One possibility, in view of certain abstract similarities between the acoustic speech and nonspeech signals (in tone/vowel-duration variation, in the suggestion that constrictions are achieved, and in gap/closure duration variation) is that the nonspeech signals might suggest a sound-producing source similar to the vocal tract in only those abstract ways. That is, the perceived producer of the nonsense signals might have a sound source that produces square-waves rather than glottal-source waves and it might have the capability of producing constrictions as the vocal tract can. If those similarities were striking enough, would the listener behave as if the same negative relation should hold between tone duration and gap duration as holds between vowel duration and closure duration in speech? The only response I could think of in order to address the question was a B/P classification.

Subjects in this group were told that they would initially be listening to some nonsense sounds each consisting of a pair of tones separated by a gap. However, these nonsense sounds had, in some sense, been modeled after the speech disyllables /apa/ and /aba/. Subjects' first task was to listen to the pair of sounds in alternation and to try to decide which sounded more like /aba/ and which more like /apa/. Shortly they would be taking tests in which they would label the /aba/-like sound with a "B" and the /apa/-like sound with a "P." Subjects were invited to listen to the demonstration sequence as many times as they wished until they had decided which pair member they would label "B" and which "P." They went on to take the 80-item test on the endpoints. They were told to label each sound either B or P, guessing if necessary. On the 100-item test, I told them that now they would be hearing not only the sounds they had been labeling but some new sounds as well. Their task was to write B if the sound they heard was more like /aba/ than like /apa/ and to write P if the sound was more like /apa/; they were instructed to guess if necessary. After a five minute break, subjects went on to take the same series of tests on the other stimulus set.

Results

A/B and L/S conditions. On the 80-item tests of the endpoints, data from two subjects were eliminated from the A/B condition, one because the subject had confused the response labels in the first 80-item test and one because the subjects' responses were random. Performance for the remaining 18 subjects averaged 90.5% correct (range 71-100%). In the 100-item test, performance on the endpoints averaged 88%, and eight subjects would not have met the 90% performance level required of their subjects by Kluender et al. (I did not apply their criterion, because it caused them to eliminate a high proportion of their subjects, half of their speech subjects and one-third of their nonspeech subjects.) However, the responses of these subjects averaged 78% and were all well above chance. In the L/S condition, performance averaged 94.0% across the two 80-item tests (range 80-100%). No subjects' data were eliminated. On the 100-item tests, performance on the endpoints averaged 93.3%, and data from three subjects failed to meet the 90% criterion of Kluender et al.[4]

Figure 6 presents the data from the 100-item tests in the A/B condition (top panel) and the L/S condition (bottom panel). Performance shows nearly the same pattern in the two panels. "A" and "short" gap classifications decrease as the gap increases in duration and there is generally a separation between the long- and short-first-tone conditions. The separation is generally opposite in direction to that predicted on the basis of auditory durational contrast, and hence it is also generally opposite to the direction of effect reported by Kluender et al. That is, a short first tone is associated with more short-gap classifications than a long first tone. This direction of difference is present numerically for every continuum member in the L/S group and for six of the ten continuum members in the A/B group. Only three continuum members in the A/B group and none in the L/S group show numerical differences in the direction reported by Kluender et al. and predicted from a durational contrast hypothesis.

An analysis of variance was performed with between-subjects factor Group (A/B, L/S) and within-subjects factors Gap duration and First-tone duration. The main effects of Gap duration and First-tone duration were significant ($F(9,306) = 167.39$, $p < .0001$; $F(1,34) = 7.93$, $p = .008$). The effect of Gap duration reflects the decrease in short-gap classifications as gap duration increases; the effect of duration reflects a larger number of short-gap classifications in the short-first-tone condition than in the long-first tone condition. The interaction of Gap duration and First tone duration was also significant ($F(9, 306) = 5.94$, $p < .001$). The interaction is significant, because the First-tone main effect holds for eight of the ten gap durations. It is slightly reversed on 90 and 100 ms gap durations, owing to the reversal of the A/B group on those stimuli.

The main effect of Group was not significant, and the only significant interaction in which Group participated was the two-way interaction with Gap duration ($F(9,306) = 2.68$, $p = .0052$). The interaction is significant because the response pattern for the L/S group is somewhat more steeply sloping than that for the A/B group as Figure 7 shows. Perhaps as a result of being told explicitly what to listen for, subjects made more short-gap classifications than the A/B group on the five shortest gap durations and fewer short-gap classifications than the A/B group on the five longest gap durations.

Finally, to assess whether the inclusion of subjects who did not meet the performance criterion of Kluender et al. (90% correct on the training test on the endpoints affected the outcome), I correlated performance on the 80-item practice test with the magnitude of the assimilation effect shown by each subject. Correlations performed separately for subjects in the A/B and L/S groups did not approach significance.

B/P group. On the 80-item tests, 10 of the eighteen subjects consistently identified the short-gap endpoint as "B" and the long-gap endpoint as "P" (87.4% on average). Seven subjects made the opposite assignment consistently (83.7%). The remaining subject made the one classification on one 80-item test and the other on the second test. A t test testing the proportion of "B" assignments to the short-gap endpoint against 50% (averaging across the two 80-item tests) was not significant ($t(17) = 1.06$, $p = .15$).

On the 100-item test, listeners responded true to their endpoint classifications. That is, subjects who labeled the short-gap endpoint "B" gave decreasing "B" classifications as gap duration increased; subjects who labeled that endpoint as "P" showed increasing "B" classifications as gap duration increased. Accordingly, the data were analysed separately for the 10 subjects consistently using the one mapping rule and for the 7 subjects consistently using the other. Both groups showed a significant effect of Gap duration ($F(9,81) = 15.95$, $p < .0001$; $F(9,63) = 4.68$, $p = .0001$), but no significant effect of first-tone duration and no interaction.

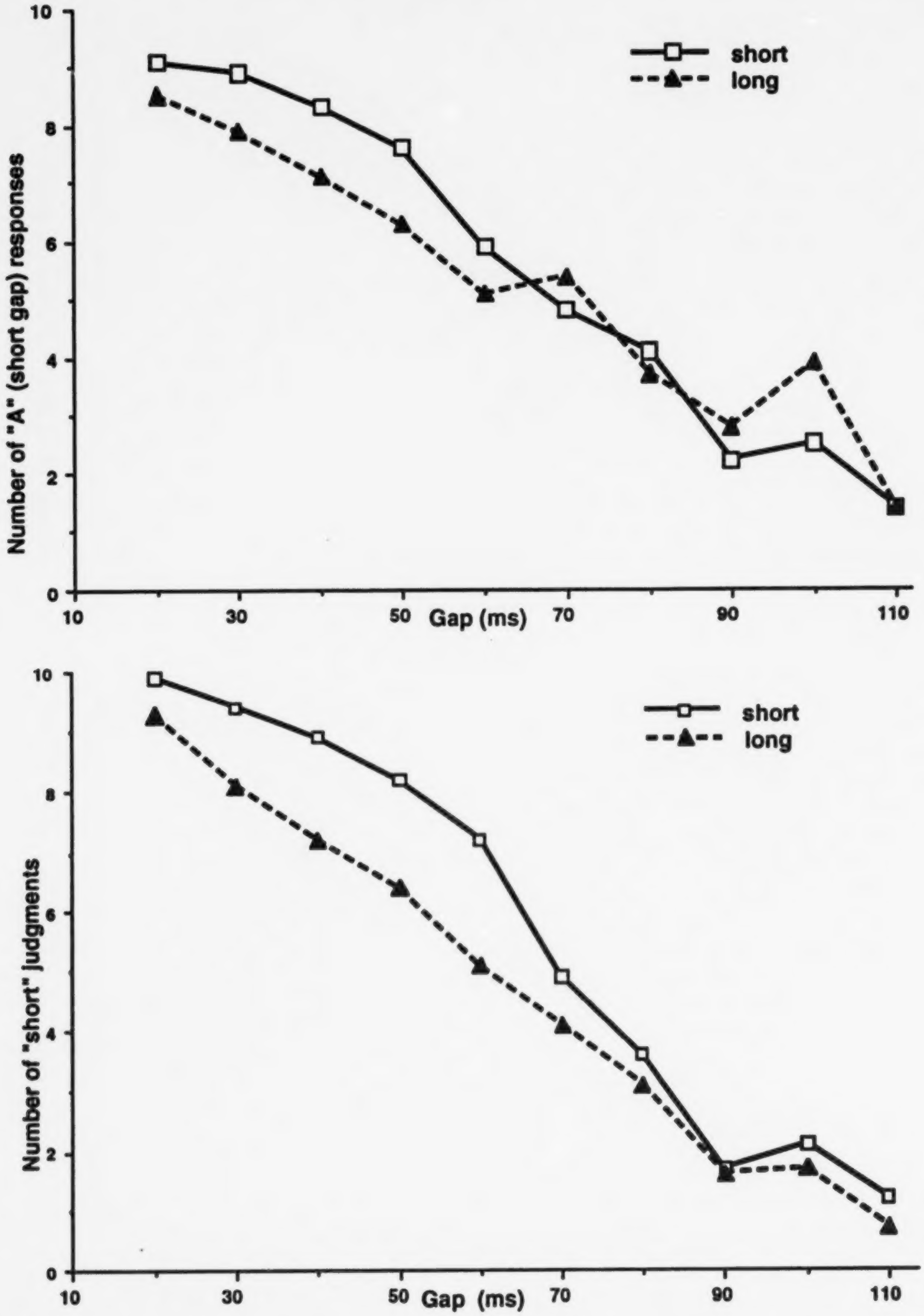


Figure 6. **Percentage of short-gap classifications in the A/B and L/S tasks of Experiment 2 as a function of first tone duration.**

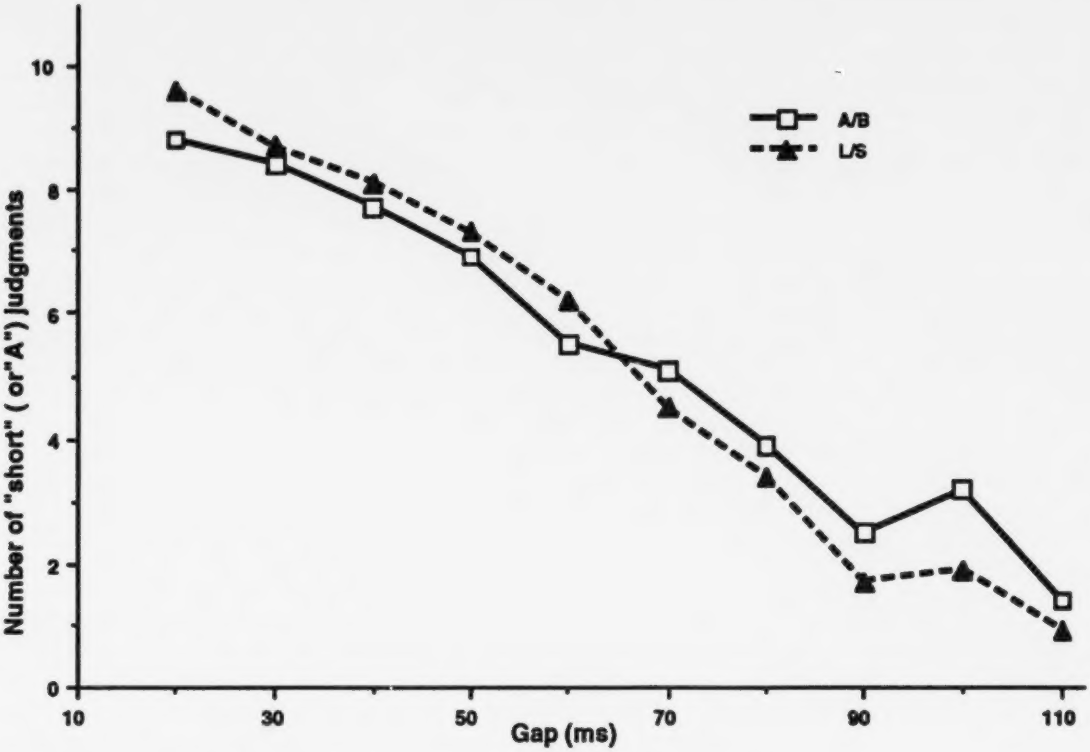


Figure 7. **Comparisons of performance levels of A/B and L/S subjects in Experiment 2.**

Discussion

Perhaps the most salient aspect of the outcome of Experiment 2 is the failure to replicate the findings of the square wave condition of Kluender et al. Those investigators reported a contrastive relation between first tone duration so that gap durations preceded by a long-duration tone were more likely to be classified with the short-gap endpoint than were the same gap durations preceded by a short tone. The findings of Experiment 2 were just the reverse even though the stimuli were modeled directly after those described by Kluender et al. For several reasons, I am confident that findings of no contrast are accurate. First, they were obtained under two distinct instructional sets, one inexplicit, following the procedures of Kluender et al., and one explicitly instructing subjects to listen for the gap duration. Second, the results replicate findings of Experiment 1 in which listeners made closure-duration (and vowel-duration) judgments of VCV syllables. The VCVs were not speech analogs of the square waves of Experiment 2, and the findings were obtained using different experimental tests than were used to obtain the data in Experiment 2. Accordingly, the finding of assimilative rather than contrastive effects have generality across a variety of stimuli and procedures. Third, the results are compatible with findings reported by Woodrow (1928) on judgments of empty intervals preceded and followed by varying duration filled intervals. Finally, Keith Kluender (personal communications, May 15, 16, 1990) has reported to me that he has run seven new subjects using his own square-wave stimuli and his procedure and now finds no effect of initial tone duration on gap-duration classifications; three of his subjects show numerical contrast, three show larger numerical assimilation effects and one shows no difference at all. While seven subjects is a small number, it is just one fewer than the number of subjects whose data were retained in the nonspeech condition of Kluender et al. These findings are similar to those of the A/B condition of the present Experiment 2. In that condition, 12 of 18 subjects showed numerical effects favoring assimilation, and the assimilation effect for this group alone is not significant if their data are analyzed separately from those of the L/S group. (However, the significant findings for the L/S group, told explicitly what to listen for, suggests that the nonsignificance for the A/B subjects and perhaps for Kluender's new subjects as well has to do with the subjects' uncertainty as to what they should be listening for.) For all of the foregoing reasons, I conclude at least that there is no durational contrast effect of filled on following open intervals in the durational ranges of stimuli in Experiments 1 and 2 and hence in the research of Kluender et al. Possibly, there is an assimilation effect.

If findings of assimilation in the present experiments were in fact to reflect auditory-system influences on perception as Kluender et al. suppose such findings to do, the influences they reflect do not explain the inverse relation between vowel duration and consonant closure in voiced and unvoiced stop-consonant contrasts that is so commonplace across languages. To the contrary, they provide an example in which language communities almost universally adopt a durational patterning that, far from exploiting distorting auditory-system influences, instead flagrantly ignores them. Looked at in this way, the findings weaken a notion that languages exploit distorting properties of the auditory system. As much as I am attracted to a conclusion such as this, I will suggest in the General Discussion why it is probably unwarranted; however, recognizing why it is involves rejecting the approach that Kluender and his colleagues adopt for examining the role of auditory system processes in speech and nonspeech perception.

GENERAL DISCUSSION

It is highly likely that the character of the human auditory system shapes the character of the sound patterns of human languages. Lindblom (e.g., 1986) has shown convincingly that it does. Language communities develop inventories of sounds that are easy to perceive and to distinguish one from the other. Moreover, analogous constraints quite generally shape the form that human artifacts take—we do not generally construct transparent floors for example, presumably because they are hard to see. The flaws in the research by Kluender et al. do not lie in the general idea that the character of the auditory system imprints itself on the form that language sound systems take but rather in the logic they use to pursue the idea.

As I have argued elsewhere (Fowler, 1990), procedures like those used in this research and used rather widely in speech research do not necessarily provide interpretable information about any auditory-system processing subserving perception of speech or of speech-like signals. The reason they do not is that they leave out of consideration an important source of patterning in listeners' responding.

The logic behind the speech-nonspeech comparisons in this and other research, I believe, is as follows. If the responses in a classification task exhibit patternings that are not obviously explainable by structure in the acoustic signal, then they are ascribed to auditory-system processes applied to the signal. This ascription is especially likely if the acoustic signals in question are nonsense signals such as the square waves of the present research. Moreover, if responses to speech stimuli pattern in the same way as responses to somewhat analogous nonsense stimuli, then both response patterns are ascribed to general auditory processes. The responses to speech signals in particular are held not to involve special-to-speech processes that presumably would impose their own distinctive patterns on responding.

In my view, the logic is mistaken. Most importantly, it fails to consider an important source of variation in responding, and that is the character of the perceived sound-producing source of the signal. Our perceptual systems have evolved to put perceivers in contact with the properties of the environment with which they interact, and that is how perceivers strive to use their perceptual systems (cf. Warren, 1981). Visual perceivers see surfaces, objects and events in the world, not the reflected light from these surfaces, objects and events. Of course it is reflected light and not the environment that stimulates the visual system; however, even so, visual systems use reflected light not as a perceptual object, but as information for the sources of its structure, and so viewers see the (visible) world. Haptic perceivers feel surfaces and objects; they use the skin deformations and joint-angle changes that felt objects cause not as perceptual objects, but as information for their causes. Indeed, the properties of objects that we both see and feel are not just compatible; they are the same properties—properties of the objects themselves, not of the perceptual systems that inform about them.

We know that, in at least some respects, the auditory system works in abstractly the same way as the haptic and visual systems in yielding environmental perceptual objects rather than percepts of the acoustic signal itself. We hear the location in space of a sound-producing object; we hear neither the location of the acoustic signals that report about the object, nor the time-of-arrival and intensity differences in the signal themselves.

No one is surprised when, in an experiment on visual perception, subjects' classification responses pattern in a way that is largely explained by patterning in the distal causes of the optical stimulation. We should not be surprised either if there is an analogous source of patterning in responses to acoustic signals. That is, the distal cause of acoustic structure is likely to affect subjects' response patterns, because that will be the listener's perceptual object. It is not necessarily the case, either, that "nonsense" acoustic signals evade that source of influence on response patterns and, therefore, directly reflect the structure in the acoustic signal itself as processed by the auditory system. Listeners certainly strive to pull information from such signals about a source in the world. In particular, they localize the sound in space, and they hear highly impoverished signals as something real (for example, as speech) given only small provocation (e.g., Remez, Rubin, Carrell, & Pisoni, 1981).

If listeners' response patterns are determined in part by properties of the perceived cause of the acoustic signal, then, when subjects' responses to speech and nonspeech signals pattern similarly, the reason for the similar pattern may as well be similarities in the perceived causes of the signals as similarities in the auditory system processes applied to the signal. I have shown (Fowler, 1990) that response patterns mimicking those of durational contrast (and of the effects of vowel duration on /b/-/w/ classifications as studied by Miller and Liberman (1979) and ascribed to durational contrast by Diehl and Walsh (1989)) can be obtained for nonspeech signals, or the reverse pattern can be obtained for similar nonspeech signals. Which outcome obtains depends on the character of the physical event that is perceived as giving rise to the signals. Here, response patterns are driven by the character of the distal sound-producing events, not by the auditory system processes applied to the signals. If response patterns to acoustic signals are to be used to draw inferences about auditory system processes, means must be developed to distinguish distal-source-produced response patterns from auditory-system-produced patterns. As far as I can tell, for example, there is no way to tell yet which source underlies the assimilation effects found in the present research.

A final comment on the logic of the research conducted to uncover a role for auditory-system processes in speech perception is that, sometimes, even where inferences to such processes seem warranted, drawing the inference does not, in itself, explain the listeners' behavior patterns. If the character of the auditory system leaves its mark on listeners response patterns to speech utterances, one possibility is that the mark reflects properties of the auditory system that have evolved to serve a function. If Warren's (1985) account of criterion shifts is apt, it is not enough or even accurate to conclude that we exhibit contrast effects because that is what the auditory system does to stimulus inputs. Rather, criterion shifts reflect a generally adaptive strategy for bring relevant past experience to bear on assessment of current perceptual experience.

REFERENCES

Behar, I., & Bevan, W. (1961). The perceived duration of auditory and visual intervals: Crossmodal comparison and interaction. *American Journal of Psychology, 74*, 17-26.

Bjork, R. (1974). The updating of human memory. In G. Bower (Ed.), *The psychology of learning and motivation, Vol 12* (pp. 235-259) New York: Academic Press.

Cathcart, E. P., & Dawson, S. (1928-1929). Persistence (2). *British Journal of Psychology, 19*, 343-356.

Davis, S., & Summers, W. V. (1989). Vowel length and closure duration in word-medial VC sequences. *Journal of Phonetics, 17*, 339-353.

Diehl, R., & Walsh, M. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America, 85*, 2154-2164.

Fowler, C. A. (1990). Sound-producing sources as objects of speech perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America 88*, 1236-1249.

Freyd, J., & Johnson, J.Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology: Learning Memory and Cognition, 13*, 259-268.

Guilford, J. P., & Park, D. G. (1931) The effect of interpolated weights upon comparative judgments. *American Journal of Psychology, 43*, 589-599.

Johnson, D. M. (1944). Generalization of a scale of values by the averaging of practice effects. *Journal of Experimental Psychology, 34*, 425-436.

Keating, P. (1985). Universal phonetics and the organization of grammars. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115-132). Orlando: Academic Press.

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics, 16*, 153-169.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. Jaeger (Ed.), *Experimental phonology* (pp. 13-44). Orlando: Academic Press.

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration and the vowel/consonant ratio in speech production. *Journal of the Acoustical Society of America, 78(6)*, 1949-1957.

Miller, J., & Liberman, A. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics, 25*, 457-465.

Parker, E. M., Diehl, R. L., & Kluender, K. R. (1986). Trading relations in speech and nonspeech. *Perception and Psychophysics, 39*, 129-142.

Pisoni, D. B., Carrell, T. D., & Gans, J. S. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception and Psychophysics, 34(4)*, 314-322.

Raphael, L. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in English. *Journal of the Acoustical Society of America, 51,* 1296-1303.

Remez, R., Rubin, P., Carrell, T., & Pisoni, D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947-950.

Repp, B., Healy, A., & Crowder, R. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 129-145.

Walker, J., Irion, A., & Gordon, D. (1981). Simple and contingent aftereffects of perceived duration in vision and audition. *Perception and Psychophysics, 29,* 475-486.

Warren, R. M. (1981). Measurement of sensory intensity. *The Behavioral and Brain Sciences, 4,* 175-189.

Warren, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review, 92,* 574-584.

Woodrow, H. (1928). Behavior with respect to short temporal stimulus-forms. *Journal of Experimental Psychology, 11(3),* 167-193.

FOOTNOTES

[*] *Journal of Phonetics,* in press.

[†] Also Dartmouth College.

[1] Two reviewers of this paper understood this argument as applying considerably more broadly than I intend it to. It is not meant to advocate collection of explicit judgments of acoustic variables in speech signals by researchers. *Au contraire* (see Fowler, 1990). Following is a clarification. Generally, to test the role of some acoustic variable in phonetic perception, it is sufficient and appropriate to cause the variable to vary and to ask subjects to make phonetic classification judgments. Generally, the interpretation of results of experiments like that in relation to the hypothesis under test is clear. That is not the case specifically for the hypothesis proposed by Kluender et al. in relation to their test and findings. It was known before their research that long vowels before an obstruent promote judgments that the consonant is voiced. Kluender et al. proposed that the effect is mediated by auditory durational contrast. To test this hypothesis, clearly, it is necessary to show that there is a contrast effect of vowel duration on judgments of consonant closure. Kluender et al. did not tell their subjects how the endpoint stimuli differed and did not tell them the basis on which they were to make their classification responses. Accordingly, if subjects heard one endpoint consonant as /b/ and the other as /p/—a not unlikely possibility—they were free to respond as if the experiment were a /b/-/p/ classification test, leaving the contrast hypothesis untested.

[2] A reviewer of this manuscript raised the concern that subjects could not do the task as instructed and, instead, made their judgments based on overall disyllable duration. However, the findings just described that subjects respond differently to a given duration difference between to-be-discriminated disyllables depending on whether the difference is in the vowel or in the closure disconfirms this idea. (Again, compare the row at zero ms difference with the column at zero ms difference.)

[3] I thank Douglas Whalen for making the square wave stimuli for me at Haskins Laboratories.

[4] Kluender et al. ran 16 subjects in their speech condition and 12 in their nonspeech condition. They eliminated half of their speech subjects and one third of their nonspeech subjects, because performance on the endpoints on the 100-item training test fell below 90%. Thus, of 28 subjects, 43% were eliminated from the experiment. Even though subjects in the present experiment did not receive feedback during training, just 31% failed to meet the 90% criterion of Kluender et al. Accordingly, I conclude that the lack of feedback during practice trials did not noticeably affect subjects' ability to learn the classification.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 141-146

P-Center Judgments are Generally Insensitive to the Instructions Given*

D. H. Whalen , André M. Cooper† and Carol A. Fowler††

The perceptual moment of occurrence of a syllable, its P-center, has frequently been examined by instructing subjects to adjust a series of speech sounds until they sounded isochronous. The present two experiments examined the effect of changing the instructions. In addition to the overall isochrony instructions, we asked subjects to align pairs of syllables so that the syllable onsets, vowel onsets, or syllable offsets sounded isochronous. In the first experiment, three of four subjects showed no difference among the first three instruction sets, and the changes introduced by the fourth went in the wrong direction. All subjects found it impossible to make alignments with respect to offsets. In the second experiment, vowel durations of two versions of some stimuli differed by 100 ms, to enhance the difference in syllable rhyme durations. Two subjects received the same instruction sets as in Experiment 1, and again found alignment with respect to offsets impossible. These subjects showed differences among the other instruction sets, although the direction and magnitude of the differences indicated that they had not succeeded in changing their timing criteria. The results indicate that P-center alignments are the syllable timing judgments that subjects most naturally make, and they may, indeed, be the only isochrony judgments that subjects can make reliably.

INTRODUCTION

The perceptual moment of occurrence of a speech sound is not its acoustic onset. Furthermore, the P-center, as it has been named, does not correspond to any single acoustic landmark in the speech signal. This is true of the acoustically defined onset (Fowler, 1979; Marcus, 1976, 1981; Morton, Marcus, & Frankish, 1976), local or global intensity peaks (Allen, 1972; Marcus 1981), the "center of gravity" (as defined by Howell, 1984, 1988, and refuted by Fowler, Whalen, & Cooper, 1988) , the number of initial consonants (Cooper, Whalen, & Fowler, 1986), or the phonetic quality of the vowel (Fox & Lehiste, 1987). Attempts to define the P-center in terms of acoustic parameters alone (Marcus 1976, 1981; Howell, 1984, 1988; Vos & Rasch, 1981) have failed (Fowler, Whalen, & Cooper, 1988). Nonetheless, some doubts remain about the best interpretation of the findings from the timing studies done to date.

One confusion has arisen from a frequently cited figure of Fowler and Tassinary (1981). In that study, speakers produced monosyllables to the beat of a metronome. The phonetic identity of the initial consonants was varied, while the syllable rhyme (in this case, /ɑd/) was the same throughout. When the syllables were displayed so that their position relative to the metronome pulse lined up, the vowel offsets happened to line up closely as well. During presentations of our work, several colleagues at Haskins Laboratories speculated that the judgments might, in fact, be

---

This research was supported by NIH Grant HD 16591, NIH Grant HD 01994 and NSF Grant BNS 8111470 to Haskins Laboratories. We thank Michael Studdert-Kennedy, Steven Marcus, and an anonymous reviewer for helpful comments.

about syllable offsets rather than onsets. Since changes in the duration of the syllable rhyme can change P-center judgments, that speculation is incorrect (Marcus, 1981; Fox & Lehiste, 1987; Cooper, Whalen & Fowler, 1988; also see Fowler and Tassinary 1981:530-531). Nevertheless, we began to wonder whether subjects did, or even could, use different alignment strategies in the method-of-adjustment technique. The present experiments were designed to see whether subjects can align syllables isochronously with respect to points other than the P-center.

Three line-up points other than the P-center were selected. The first was the acoustic onset of the syllable. Perhaps if subjects are instructed to line them up, they could do so, regardless of the rest of the syllable. The second was the vowel onset, which is the acoustic landmark whose variation in timing with respect to the onset of the preceding consonants is most like that of the P-center. The third, already discussed, was the syllable offset.

If the different sets of instructions given to subjects result in different timing alignments, then the P-center literature must be interpreted cautiously. Not only would the exact instructions given to subjects be seen as crucial for determining the correct interpretation of the results, but it would be clear that timing patterns other than P-center alignment could have been studied. Alternatively, if the different instructions do not affect subjects' timing judgments, then the interpretations of the previous literature would be bolstered. Not only would the exact nature of the instructions be irrelevant, but the notion that the P-center is influenced by the entirety of the speech signal (Marcus, 1981; Cooper, Whalen & Fowler, 1986) would receive further support.

EXPERIMENT 1

Subjects were asked to align syllable pairs so that the whole syllable or one of three acoustic events was isochronous. The alignments were made by modifying the intervals between two syllables in continuous alternation.

Method

Stimuli. The syllables used were /bɑd/, /tɑd/, and /ʃɑd/. These were selected because they have three distinct P-centers, when measured from the acoustic onset (Fowler & Tassinary, 1981). Several examples of each syllable were recorded on audio tape. These were then digitized (10 kHz sampling rate, 12 bit resolution) and one token of each was selected for use in the experiment. The durations were 414, 414, and 549 ms for /bɑd/, /tɑd/, and /ʃɑd/ respectively. The durations of the voiceless segments at the beginning of the syllable (either burst plus aspiration or fricative noise) were 13, 66, and 179 ms, respectively.

Subjects. Three of the subjects (AMC, CAF, and MRS) had participated in other timing studies (Cooper, Whalen & Fowler, 1986, 1988). The fourth subject (HE) was a young male volunteer, who was paid for his services. The first two subjects were two of the authors, and the last two were naive to the purpose of the experiment. All subjects were native speakers of some variety of American English, though HE had spent his childhood in Jamaica.

Equipment and Procedure. Each syllable was paired with each of the other two, giving six permutations. For each pair, the first syllable presented was considered the standard, and the other syllable's timing could be changed relative to it. At the beginning of each trial, there was a 50 ms gap between the offset of the first syllable and the onset of the second; the interval between the offset of the second syllable and the onset of the next repetition of the first was the longest possible inter-stimulus-interval (ISI). The subjects would then modify this ISI in accordance with the instructions given.

Two sets of equipment were used. Each involved computer control of the ISI, but one used keyboard presses to change the ISI (used by CAF and MRS), and the other used an analog potentiometer (for AMC and HE). The other difference between the two setups was in the interval between the onset of like syllables. For the first system this interval was the duration of those syllables plus 550 ms. For the second system, there was a constant 1400 ms from the onset of the fixed syllable to the next repetition of this syllable.

Four blocks, each with a different set of instructions, were run for each subject. The instructions were to adjust the timing of the syllables until: 1) the whole syllable sounded isochronous as if timed to the beat of a metronome, 2) their onsets were isochronous, 3) their vowel onsets were isochronous, or 4) their offsets were isochronous. All subjects received the blocks of instructions in this order. The syllable pairs were presented in random order. Subjects could take as long as they needed to make their adjustments. There were four repetitions of each of the six pairings for CAF and MRS, and six for AMC and HE.

Results

The manipulation of syllable order within a trial (e.g., /bɑd/-/ʃɑd/ versus /ʃɑd/-/bɑd/) did not affect the results, so all measurements were adjusted to ignore this factor. Each alignment value reflects the divergence away from the midpoint of the /bɑd/-to-/bɑd/ interval (or the /tɑd/-to-/tɑd/ interval for trials without /bɑd/ stimuli). Positive numbers show that the /tɑd/ or /ʃɑd/ syllable's onset was earlier than this midpoint, reflecting agreement with P-center measurements made in the past.

Statistical analyses were performed only on the first three instruction sets. None of the four subjects completed the offset alignment task, claiming that they had no idea what they were doing. We suspected that the extreme difficulty of the task might have been due to the lack of variation in the durations of the syllable rhymes, so we did not interpret this refusal further.

The mean measurements for the three pairings in the three instruction sets are shown in Table 1. An analysis of variance with the factors Instruction Set, Subject, and Pairing was run on the individual trial values. There was no main effect of Instruction Set ($F(2,276) < 1$); however, effects of Subject and Pair were significant ($F(3,276) = 7.67$, $F(2,276) = 34.96$, $p < .001$, respectively). Interactions between Instruction Set and Subject, and between Subject and Pair were also significant ($F(6,276) = 2.89$, $F(6,276) = 8.70$, $p < .01$ for the two respectively). Individual subject analyses showed that subject HE had a significant effect of Instruction Set ($F(2,63) = 3.46$, $p < .05$). However, his means for the acoustic onset condition were larger than for the P-center and vowel onset conditions (144 versus 83 and 89 ms), indicating that he introduced *more* deviation from acoustic isochrony when he attempted to align acoustic onsets.

Table 1. *Mean timing alignments for four subjects in three instruction conditions, Experiment 1. Numbers in parentheses are the values that should have been obtained if the subjects had been able to do the task.*

		Instruction Set	
Pair	P-center	Acoustic	Vowel
		Onset	Onset
b-ʃ	122	131 (0)	124 (166)
b-t	42	39 (0)	33 (53)
t-ʃ	75	95 (0)	73 (113)

As a final check to ensure that the amount of variability in the data was not so extreme as to obscure a difference between the P-center and acoustic onset conditions, we modeled perfect performance on the acoustic onset condition by subtracting from each value the mean for that pair for that subject. Thus the means for that condition came out to zero, which would have indicated perfect alignment of the acoustic onsets. If that had been the case, then the instruction set would have had a significant effect ($F(1,168) = 50.11$, $p < .001$). Thus if there had been a sizable effect, this amount of data would have discovered it. The one subject who showed an effect showed it in the wrong direction.

Discussion

None of the subjects succeeded at all in aligning the syllables with respect to acoustic onsets. All but one subject made statistically indistinguishable temporal alignments for each of the other three instruction sets. In fact, the one subject who showed a difference in the instruction sets tended to push the syllables farther from acoustic onset isochrony in his "acoustic onset" alignments than in his P-center alignments. Even the vowel onsets were more accurately aligned (though not significantly more) in the P-center instruction condition than in the vowel onset condition. In short, no matter what the instructions were, subjects generally seemed to base their timing judgments on the P-center.

EXPERIMENT 2

In Experiment 1, subjects may have found the syllable offset alignment task difficult to perform because correct alignment of the offsets would make the P-centers very close as well. Thus it may have been difficult to ignore the more robust P-center alignment and focus in on the more difficult offset alignment, despite the instructions. In Experiment 2, larger variations in the duration of the vocalic segment were introduced into some of the syllables in an effort to make the syllable offset instructions easier to carry out.

Method

Stimuli. Several tokens of /bɑ/, /tɑ/, and /ʃɑ/ were recorded on audio tape and digitized as described for Experiment 1. One token of each type was chosen such that all the vocalic segments were roughly equal in intensity and fundamental frequency. The original durations were 414, 412, and 538 ms for /bɑ/, /tɑ/ and /ʃɑ/ respectively. The onset-to-vocalic-segment durations were 27, 79, and 157 ms for /bɑ/, /tɑ/ and /ʃɑ/ respectively. Two additional stimuli were made from /bɑ/ and /ʃɑ /. A shortened version was made by deleting whole pitch periods until approximately 50 ms had been removed. A lengthened version was made by repeating the same pitch periods that had been removed from the shortened versions. Thus shortened /bɑ / was 364 ms in duration, and lengthened /bɑ / was 464 ms. Shortened /ʃɑ / was 486 ms, and lengthened /ʃɑ / was 590 ms.

Subjects. Two subjects participated in this experiment. They were AMC, who had participated in the previous experiment, and a young male research assistant (RFP), who had participated in a previous study (Cooper, Whalen and Fowler, 1986) but was naive to the purpose of the present study. Both were native speakers of American English.

Equipment and Procedure. The setup with the keyboard response was used. For this experiment, however, all of the stimuli were adjusted within a 1400 ms window regardless of the duration of the stimuli.

To reduce the number of judgments that had to be made, only some combinations of stimuli were used. The shortened and lengthened versions of /bɑ / were paired with original /tɑ/ and /ʃɑ/, to give four of the pairs. The shortened and lengthened versions of /ʃɑ/ were paired with original /tɑ/, to give the other two. The same instructions were given as in Experiment 1. AMC gave six judgments on each pair with each syllable starting the series, and RFP gave eight.

Results

The greater variation in the stimulus durations in this experiment was not enough to make the subjects confident at all in their offset alignments. Subjects had such difficulty in performing the offset alignment task that they eventually gave up. One subject (AMC) went through much of this condition, but quit when he discovered that his last three answers had been ones which allowed the minimum (50 ms) of silence between one syllable and the next, i.e., the equivalent of the starting position. No data will be reported from this instruction set.

The measurements were transformed as in Experiment 1, with positive numbers indicating that /tɑ/ and /ʃɑ/ had acoustic onsets earlier than /bɑ/. For the manipulation of vowel duration, this had the unfortunate effect that the predicted difference went in one direction for /bɑ/ and in the other direction for /ʃɑ/. Thus the effect of shortening or lengthening the vocalic segment shows up first as an interaction between the Duration factor and the Pair factor.

The results are shown in Table 2. An analysis of variance was performed on the individual trial values with Instruction Set, Subject, Pair and Duration as the factors. The first three main effects were significant ($F(2,468) = 11.46, p < .001$; $F(1,468) = 7.14, p < .01$; $F(2,468) = 378.55, p < .001$ respectively). As expected, Duration was not significant as a main effect ($F(1,468) = 0.85$) but only as an interaction with Pair ($F(1,468) = 14.15, p < .001$). There were further interactions of Instruction Set and Subject, ($F(2,468) = 9.65, p < .001$), Instruction Set and Pair, ($F(2,468) = 8.49, p < .001$), and Subject and Pair ($F(2,468) = 3.83, p < .05$).

Table 2. *Mean timing alignments for two subjects in three instruction conditions, Experiment 2. Numbers in parentheses are the values that should have been obtained if the subjects had been able to do the task.*

	Instruction Set		
Pair	P-center	Acoustic Onset	Vowel Onset
b(shortened)- ʃ	156	103 (0)	137 (130)
b(lengthened)- ʃ	133	105 (0)	122 (130)
b(shortened)-t	22	26 (0)	14 (52)
b(lengthened)-t	13	28 (0)	4 (52)
t- ʃ(shortened)	106	81 (0)	92 (78)
t- ʃ(lengthened)	132	104 (0)	126 (78)

Tests of each of the three pairings of the instruction sets revealed that each was distinct from the other, with grand means of 93, 74 and 82 ms for P-center, Acoustic Onset and Vowel Onset instructions respectively. The acoustic onset adjustments were generally closer to physical isochrony than the P-center adjustments were. However, the acoustic adjustments were still much closer to the P-center values than they were to physical isochrony. The vowel onset values were close to the physical target of 87 ms. However, other aspects of the data complicate the picture that these mean values depict. The effect of syllable duration on the isochrony judgments was roughly equivalent for the P-center and Vowel onset instructions. If the subjects had succeeded in aligning the syllables by the vowel onset, the duration of the vocalic segment should not have had any effect on the timing judgments. Further, a test of the variances in the syllable onset and vowel onset conditions revealed that P-center alignments accounted for the data better than the physical targets for those conditions. For each condition, we generated two sets of variances. One was based on the P-center alignment for that cell, the other, on the physical target value for the condition. These physical target values were 0 ms for the syllable onset condition and, in the vowel onset condition, 130, 52 and 78 ms for /bɑ/-/ʃɑ/, /bɑ/-/tɑ/, and /tɑ/-/ʃɑ/ respectively. A test of the absolute values of these differences showed that the judgments were less different from the P-center alignments than they were from the physical targets in both the syllable onset (39 ms vs. 80 ms, $F(1,156) = 73.20$, $p < .001$) and the vowel onset conditions (29 ms vs. 39 ms, $F(1,156) = 37.23$, $p < .001$). Whatever it was that subjects were doing differently, it was not likely to have been finding the syllable or vowel onset. Though the onset instructions elicited judgments significantly different from the P-center instructions, these judgments were still significantly closer to P-center values than to their respective physical targets. Further, there were some cases where the vowel onset judgments were closer to the acoustic onset physical target than the vowel onset physical target (/bɑ/-/tɑ/), and vice versa (/tɑ/-/ʃɑ/).

Discussion

The fact that subjects were unable to align syllables with respect to vowel offsets is a strong indication that they did not do so in earlier experiments. This is supported by the fact that the 100 ms change in the duration of the stimuli in this experiment resulted in an overall shift of 15 ms in the timing judgments, rather than none or 100 ms. This is consistent with the small effect of changes in the vocalic segment on the P-center obtained in other research (cf. Marcus, 1976, 1981; Cooper, Whalen & Fowler, 1988).

Although the subjects in this experiment succeeded in differentiating the three remaining instruction sets, this result does not suggest that subjects were able to align the syllables with respect to acoustic onsets or vowel onsets. That is, the fact that something was done differently did not imply necessarily adjusting the syllables in accordance with the instructions.

CONCLUSIONS

Timing judgments for speech stimuli are generally insensitive to the instructions given. Subjects tend to be unable to adopt a specific alignment strategy other than the one they adopt spontaneously. In the first experiment, we found that only one subject showed any differences in alignment under three sets of instructions, and that was in the wrong direction. In the second experiment, subjects did make distinctions among the instruction sets, but only occasionally in the direction that was appropriate. In both experiments, subjects could not align syllable offsets. We are thus led to two conclusions:

1) The previous literature on speech timing, especially that using the method of adjustment, is indeed about the timing of P-centers, regardless of the exact nature of the instructions that were given. A corollary of this is that the close alignment of the acoustic offsets in Figure 1 of Fowler and Tassinary (1981) is an artifact of the close match in syllable rhyme durations used there.

2) Not only are there no obvious acoustic markers for the P-center in speech, listeners are unable to make conscious timing judgments based on the acoustic markers of syllable onset, vowel onset, and syllable offset. Instead, the P-center seems to be the only perceptual event at which syllables can be aligned to isochrony.

REFERENCES

Allen, G. (1972). The location of rhythmic stress beats in English: An experimental study I. *Language and Speech, 15,* 72-100.

Cooper, A. M., Whalen, D. H., & Fowler, C. A. (1986). P-centers are unaffected by phonetic categorization. *Perception & Psychophysics, 39,* 187-196.

Cooper, A. M., Whalen, D. H., & Fowler, C. A. (1988). The syllable's rhyme affects its P-center as a unit. *Journal of Phonetics, 16,* 231-241.

Fowler, C. A. (1979). "Perceptual centers" in speech production and perception. *Perception & Psychophysics, 25,* 375-388.

Fowler, C. A., & Tassinary, L. G. (1981). Natural measurement criteria for speech: The anisochrony illusion. In J. Long and A. Baddeley (Eds.), *Attention and performance* (Vol. 9, pp. 521-535). Hillsdale, NJ:Erlbaum.

Fowler, C. A., Whalen, D. H., & Cooper, A. M. (1988). Perceived timing is produced timing: a reply to Howell. *Perception & Psychophysics, 43,* 94-98.

Fox, R. A., & Lehiste, I. (1987). The effect of vowel quality variations on stress-beat location. *Journal of Phonetics, 15,* 1-13.

Howell, P. (1984). An acoustic determinant of perceived and produced anisochrony. In M. P. R. Van den Broecke and A. Cohen (Eds.), *Proceedings of the 10th International Congress of*

Phonetic Sciences (pp. 429-433). Dordrecht, The Netherlands: Foris.

Howell, P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope: I. *Perception & Psychophysics, 43,* 90-93.

Marcus, S. M. (1976). *Perceptual centres.* Unpublished doctoral dissertation, Cambridge University.

Marcus, S. M. (1981). Acoustic determinants of Perceptual Center (P-center) location. *Perception & Psychophysics, 30,* 247-256.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual Centers (P-centers). *Psychological Review, 83,* 405-408.

Vos, J., & Rasch, R. (1981). The perceptual onset of musical tones. *Perception & Psychophysics, 29,* 323-335.

FOOTNOTES

[*] *Phonetica, 46,* 197-203 (1989).

[†] Also Yale University. Now at University of Michigan.

[††] Also Dartmouth College.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 147-170

Perceptual Restoration of a "Missing" Speech Sound: Auditory Induction or Illusion?*

Bruno H. Repp

This study investigated whether the apparent completeness of the acoustic speech signal during phonemic restoration derives from a process of auditory induction (Warren, 1984) or segregation, or whether it is an auditory illusion accompanying the completion of an abstract phonological representation. Specifically, five experiments tested the prediction of the auditory induction (segregation) hypothesis that active perceptual restoration of an [s] noise that has been replaced with an extraneous noise would use up a portion of that noise's high-frequency energy and consequently change the perceived pitch (timbre, brightness) of the extraneous noise. Listeners were required to compare the pitch of a target noise, which replaced a fricative noise in a sentence, with that of a probe noise preceding or following the speech. The first two experiments found a significant tendency in favor of the auditory induction hypothesis, although the effect was small and may have been caused by variations in acoustic context. The following three experiments, which used a larger variety of stimuli and controlled context more carefully, yielded negative results. Phoneme identification results collected in the same experiments, as well as informal observations about the quality of the restored phoneme, suggested that restoration of a fricative phone distinct from the extraneous noise did not occur; rather, the spectrum of the extraneous noise itself influenced phoneme identification. These results suggest that the apparent auditory restoration which accompanies phonemic restoration is illusory, and that the schema-guided process of phoneme restoration does not interact with auditory processing.

INTRODUCTION

Phoneme restoration: Two hypotheses

One of the best-known and most interesting findings in speech perception research is the "phonemic restoration illusion" first demonstrated by Warren (1970; Warren & Obusek, 1971; Warren & Sherman, 1974). In Warren's original experiments, subjects listened to a sentence in which the acoustic signal pertaining to one phoneme, the fricative /s/, had been excised and replaced with an extraneous sound or with silence. When an extraneous sound, such as a cough, was present, listeners claimed to hear the speech as intact and could not localize the cough (or the speech sound masked by it) accurately within the sentence. The silence, however, was usually heard as replacing the /s/. Warren (1976, 1984) interpreted these findings as reflecting general principles of auditory perception: "Phonemic restoration can be considered as a verbal form of heterophonic auditory induction, and if the neural units stimulated by the extraneous sound include those which would be stimulated by the missing phoneme (as identified through contextual cues), then perceptual synthesis of this phoneme may occur" (1976, p. 403). This perceptual synthesis, however, is "governed by linguistic rules" (1984, p. 375).

Later research by Samuel (1981a, 1981b) confirmed that the extraneous sound must have some spectral resemblance to the missing speech sound for phonemic restoration to be effective. Samuel's innovative contribution was methodological:

This research was supported by NICHD Grant HD-01994 and by BRSG Grant RR-05596 to Haskins Laboratories. I am grateful to Diana Matson, Sal Miranda, and Lisa Zsiga for assistance, and to Carol Fowler, Arthur Samuel, Richard Warren, and D. H. Whalen for helpful comments on earlier drafts of this manuscript. A short report of Experiments 1 and 2 appeared previously in Repp (1989).

Instead of relying merely on subjects' inability to identify the "missing" sound, he employed a one-interval discrimination paradigm in which subjects had to distinguish between speech stimuli in which the extraneous sound *replaced* part of the speech signal, and stimuli in which the extraneous sound was merely *added* to the critical speech segment. The relative difficulty of this discrimination was considered a measure of the *strength* of phonemic restoration, the reasoning being that "replaced" stimuli cause auditory synthesis of the missing phoneme and therefore should sound just like "added" stimuli if restoration is complete. The discriminability and bias indices of signal detection theory (d' and Beta) were the dependent variables in a number of subsequent studies that investigated various conditions that might influence the strength of the phonemic restoration effect (Samuel, 1981a, 1981b, 1987, in press; Samuel & Ressler, 1986; Trout & Poser, 1990).

If a "replaced" stretch of the speech signal appears to be present in the listeners' subjective auditory experience, where does the auditory information for that restoration come from? One possibility is that the restoration is an auditory illusion that arises from context-induced *phonological* completion and consequent phonetic imagery (i.e., inner speech). That is, perhaps subjects do not (or cannot) actually make judgments about the intactness of the auditory input signal, and instead report on an abstract phonological or imaginary phonetic representation that is activated in the process of word recognition (cf. Foss & Blank, 1980). In that case, the extraneous sound replacing the missing phoneme provides "bottom-up confirmation" (Samuel, 1981b), presumably by facilitating lexical access to the extent that it *masks the absence* of the missing phoneme, but otherwise it does not participate in the restoration process. The strength of the illusion should then depend on the success and speed of lexical access, and this hypothesis is indeed supported by Samuel's findings that the illusion is reduced for nonwords (Samuel, 1981a) and lexically ambiguous words (Samuel, 1987), whereas it is enhanced by pre-exposure to nonwords (Samuel, 1981a). It is further supported by the result (Samuel and Ressler, 1986; Samuel, in press) that cuing subjects' attention to the critical phoneme in the word reduces the illusion, which suggests that the speech input is not perceived as complete when its auditory memory trace is examined closely. However, as is well known from categorical perception research (see Repp, 1984), access to detailed auditory properties of speech is quite difficult in high-uncertainty listening situations, so the absence of genuine auditory restoration may go unnoticed when the phonological structure is complete.

There is another possibility, however, which becomes evident when the "added" condition is considered, where the original acoustic information supporting perception of the critical phoneme is actually present in the input signal. In that condition, it seems plausible that the listener's auditory system would *segregate* the extraneous sound from the speech, so that the speech signal is heard as intact because it *is* intact. Since the auditory system does not "know" in advance whether the original speech information is present or absent, one might suppose that it would attempt the same, context-guided segregation procedure in the "replaced" condition. Since the extraneous sound must be spectrally similar to (i.e., a potential masker of) the missing speech sound for restoration to occur (Warren, 1984; Samuel, 1981a, 1981b), it may be argued that the replaced speech sound is not really missing; rather, it is *hidden* in the extraneous sound and must be extracted and separated from it. This separation may be less effective or complete than in the case of "added" stimuli—only some resemblance of the missing speech sound may be restored—but the basic process may be the same. According to this reasoning, top-down expectations guide the listeners as to what to extract, but these expectations interact with primitive auditory processes that attempt to reconstitute the speech signal. Phonemic restoration is not an illusion in that case; it is an instance of auditory source segregation, and the acoustic signal does not simply confirm a restoration accomplished at a higher level but provides the material out of which the restored sound is carved by the auditory system. This is essentially Warren's (1984) "auditory induction" account.

These two possibilities have implications not only for the perception of the restored speech segment, but also for the perception of the extraneous sound. According to the first hypothesis (henceforth, the *top-down completion hypothesis*), the restored speech should "sound" quite normal because subjects report on an ideal internal image accompanying the restored phoneme, and the extraneous sound should sound very similar to the way it sounds when it occurs in isolation. According to the second hypothesis (the *segregation hypothesis*), the extraneous sound is partitioned into a likeness of the acoustic speech segment expected at that point and into a *residue* that is perceived as an extraneous sound. Because the acoustic energy necessary to support perception of the speech segment is subtracted from the extraneous sound, as it were, the perceived sound quality (as well as the loudness) of the extraneous sound should change when restoration occurs. Likewise, the sound of the restored speech segment may not be fully natural if the extraneous sound provides only a partial "spectral blanket." Because listeners find it generally very difficult to judge the auditory quality of speech segments, particularly in a sentence context, the present study focused on the predictions concerning the perception of the extraneous sound.

In addition to Warren's and Samuel's work, two other contemporary theories are directly relevant to the hypotheses contrasted here.[1] In his thoughtful and comprehensive discussion of auditory scene analysis, Bregman (1990) distinguishes between primitive and schema-driven processes. Phonemic restoration is clearly schema-driven: The acoustic-phonetic context alone is hardly sufficient to suggest the missing speech segment to a listener. Linguistic knowledge, or a set of speech schemata in Bregman's terminology, is needed to infer the identity of the missing segment from the context. The question posed here is whether this schema-driven process interacts with primitive auditory processes of source assignment and timbre perception. Bregman believes that it does not: "Speech schemas are the same as other recognition schemas in the following way: when they make use of the information that they need from a mixture, they do not remove it from the array of information that other description-building processes can use" (p. 638). And, with specific reference to phonemic restoration, he states that it is "based on schema-governed stream segregation and for this reason will not give rise to a distinct residual" (p. 372). According to his theory, a residue-forming process of auditory stream segregation would occur only when the segregation is directly induced by the surrounding acoustic context, as demonstrated in spectrally complex but steady-state nonspeech sounds whose spectral components may be "captured" by a preceding and following simpler sound according to the law of continuity (e.g., Warren, Obusek, & Ackroff, 1972; Bregman & Pinker, 1978; Bregman, 1990). Thus, Bregman's view, like Samuel's, seems to be closer to the top-down completion hypothesis than to the segregation hypothesis.

In contrast, Mattingly and Liberman (1988, 1990; Liberman & Mattingly, 1989) have proposed that the speech perception system "pre-empts" auditory scene analysis, taking what it needs from the auditory signal and leaving a residue for general auditory processes to deal with. This hypothesis is based primarily on evidence from the duplex perception paradigm, which has been investigated intensively by Liberman and his colleagues (see, e.g., Mann & Liberman, 1983; Whalen & Liberman, 1987). There are some interesting parallels between duplex perception and phonemic restoration. The dichotic duplex perception paradigm juxtaposes an ambiguous syllable lacking a critical formant transition (the "base") in one ear with the isolated formant transition in the other ear. Listeners report hearing the intact syllable as well as the chirp-like isolated transition, just as they claim to hear both the intact speech and the extraneous sound in the phonemic restoration paradigm. In duplex perception, as in phonemic restoration, the question then arises whether the full syllable is constructed at a higher level (by combining phonetic elements perceived as coming from separate sources, in this instance), so that its perceived auditory presence in lieu of the base is illusory (cf. Nusbaum, Schwab, & Sawusch, 1983), or whether the missing transition is actively segregated from the spectral information entering the opposite ear, at an early stage in binaural processing.

Whalen and Liberman (1987) clearly favor the latter interpretation and have provided some empirical support for it in a monaural duplex perception situation: By varying the relative amplitude of a sinusoidal frequency transition embedded in a syllable, they determined that the detection threshold for the transition as a whistle was much higher than the phoneme identification threshold. That is, the transition contributed to speech perception even when it could not be heard as a separate auditory event. A similar finding for the standard dichotic duplex perception paradigm was reported by Bentin and Mann (1983, 1990). On the basis of these results, Whalen and Liberman (1987) and Mattingly and Liberman (1988) suggested that the speech system takes the auditory information it needs and leaves a less intense frequency transition for the general auditory system, too weak to be detected unless its intensity is raised substantially. Applied to the phonemic restoration paradigm, this view might predict that the speech system will grab the information it needs from the acoustic spectrum provided by the extraneous sound, leaving an impoverished residue. This speech-specific "preemptiveness hypothesis" thus coincides with the (more generally schema-based) segregation hypothesis, proposed above.[2]

Phone restoration: Methodological considerations

The present study is about *phone restoration*, more than about phoneme restoration. The *phoneme*, an abstract phonological category without a definite sound, is distinguished here from the *phone*, a concrete auditory percept based on a relatively homogeneous acoustic speech segment and having elementary auditory properties such as pitch, timbre, and duration (see Repp, 1981a). Whereas, in the phoneme restoration paradigm, all acoustic cues for a given phoneme are excised from the acoustic signal, in the present study only one acoustic segment was removed, specifically the noise portion of a fricative consonant. The approach taken here required that the critical segment be acoustically as simple as possible, so that the effects of its segregation from an equally simple extraneous sound could be predicted, perceived, and measured. Other cues to fricative identity—viz., the vocalic formant transitions preceding and following the noise portion—remained intact and, without any doubt, enhanced restoration of the correct fricative phoneme (cf. Mann & Repp, 1980; Mann & Soli, 1991). Since fricative phoneme restoration has been demonstrated numerous times in speech signals harboring no residual fricative cues, the occurrence of phoneme restoration in the presence of such cues was taken for granted.

Given that there are residual fricative cues in the surrounding speech signal, auditory restoration of a fricative noise (if it occurs) may be guided not only "from above" by a phonetic speech schema but also "from below" by local acoustic properties that call for a "good continuation" through the gap covered up by the extraneous noise. Soli (1981), for example, has demonstrated that formant transitions extend into and out of fricative noises and thus provide acoustic continuity with the surrounding vocalic segments. There may even be low-amplitude noise or breathiness in the immediately adjacent vocalic segments, due to the change in laryngeal action between voiced and voiceless segments, which may induce restoration of a noisy speech segment. Although the acoustic context alone cannot tell the auditory system the precise spectrum and amplitude of the missing phone (information that a speech schema might provide), it does provide pointers to some acoustic properties of that noise. Thus, conditions in the present experiments were maximally conducive to auditory restoration and segregation of the "missing" phone: Given the guidance provided by the acoustic context, both at the level of primitive auditory scene analysis and indirectly via the top-down influence of lexical/phonological units stimulated by the residual cues, the restored fricative noise should be partialled out from whatever sound occurs in its place, if this is at all what happens in phoneme restoration. Only if phoneme restoration is a pure top-down process (i.e., never "consumes" acoustic energy) would the experimental results be expected to favor the top-down completion hypothesis.

The experimental task required the subjects to pay attention to the *extraneous sound* (a noise) which replaced the target phone (typically [s]) in a brief sentence and to compare its auditory quality to that of a probe noise occurring either before or after the sentence. According to the top-down completion hypothesis, the target noise should be perceived veridically: If the target and probe noises are physically identical, they should sound the same. According to the segregation hypothesis, however, physically identical target and probe noises should not be perceived as identical. The nature and direction of the perceived difference should reflect the acoustic properties of the restored speech sound: Since an [s] noise is characterized by strong high-frequency energy, the residue of the target noise should be depleted of energy in that region. Noises can be described as having a pitch-like quality or relative "brightness" of timbre that depends on their spectral composition (see, e.g., Glave, 1973; Hesse, 1982). An [s] noise has a relatively high pitch or bright timbre, which reflects the predominance of high-frequency components mostly above 4 kHz, whereas white noise has a somewhat lower (indefinite) pitch or darker timbre, which reflects its flat spectrum, and low-pass filtered noise has an even lower pitch. Therefore, subjects can be asked to make judgments about the relative pitch height of pairs of noises, and the term "pitch" will be used henceforth to refer to the subjective dimension along which the noises were compared. The prediction of the segregation hypothesis was, then, that the target noise should be perceived as having a lower pitch than the probe noise.

One possible complicating factor was that the acoustic context surrounding the target noise might have an influence on the perceived pitch of the noise for reasons that are unrelated to the hypothetical process of phone restoration. For example, the following context might interfere with a listener's memory for the noise pitch in specific ways, and both preceding and following context may exert auditory assimilation or contrast effects on perception of the noise spectrum. The memory problem was addressed by having the probe noise either precede or follow the speech in which the target noise was embedded. If the acoustic speech signal interfered with memory for the target noise, then this interference should be reduced when the probe noise comes first and subjects make a decision immediately upon hearing the target noise. The other problem (i.e., possible assimilation or contrast effects) required some control stimuli in which there was similar acoustic context but no incentive for fricative restoration. In Experiments 1 and 2, stimuli conducive to restoration of /s/ were compared with stimuli containing a /t/ instead. The phone replaced by the extraneous noise in the case of /t/ was silence (the stop closure interval), so that *no acoustic energy* was to be restored for perception of /t/. This seemed like an appropriate control condition to start with.

EXPERIMENT 1

According to the top-down completion hypothesis, a white noise burst that replaces an [s] noise in a sentence should be perceived as having the same pitch as an identical white noise burst occurring before or after the sentence (if we disregard the possibility of auditory context effects for the moment). According to the segregation hypothesis, however, the white noise in the sentence should sound *lower* in pitch (and also less loud, though this was not tested here) than a white noise probe because the [s] frication has been perceptually subtracted from it, as it were, leaving a residual noise in which frequencies below 4 kHz predominate. It follows, then, that there must be some noise with relatively low-frequency predominance that, when employed as a probe, is perceived as matching the white noise target. The precise spectrum of that probe noise is not obvious, however, since it is not clear "how much" [s] noise the auditory system needs to subtract from the white noise inside the sentence to satisfy the needs of the speech processor. A strong version of the segregation hypothesis would predict that something close to the original [s] noise is subtracted, based on expectations derived from the speaker's voice and knowledge about speech in general. However, a much weaker copy of that noise or a noise with a somewhat distorted spectrum might suffice for /s/ phoneme restoration.

To construct a noise probe that, according to the strong segregation hypothesis, should match the white noise target, the original [s] noise somehow needed to be subtracted physically from the white noise burst. Since arithmetic subtraction of the digitized [s] noise waveform from the white noise waveform does not have the desired spectral consequences (with random noise, waveform subtraction is essentially equivalent to waveform addition), a low-pass filtered noise (N) was instead created that, when the [s] noise waveform was added to it, resulted in a noise with an approximately flat spectrum (NS). N was the noise probe that was expected to match the NS target: If listeners restore [s] fully by segregation from NS, they should hear the intact speech accompanied by the residual extraneous noise N. (Note that the situation can be described either as NS replacing [s], or as N being added to [s]!) To simulate the more typical "added" condition, another noise was constructed by adding the waveforms of NS and [s] (NSS). According to the segregation hypothesis, NSS replacing [s] in a sentence should be perceived as [s] accompanied by NS. According to the top-down completion hypothesis, however, NS and NSS targets replacing [s] should be perceived as matching NS and NSS (not N and NS) probes, respectively.

In Experiment 1, then, three noise probes (N, NS, and NSS) were paired with two noise targets (NS and NSS). The target noise replacing [s] occurred either at the beginning or at the end of a trisyllabic word embedded in the constant carrier phrase "Say ... again," and the noise probe occurred either before or after the sentence. The temporal locations of the target and probe noises were varied to examine the role of auditory memory in the task. Because of limits on the size of the design, only one word of each type was employed: "seminar" and "happiness." The final /s/ of "happiness" is predicted by the preceding context, whereas the initial /s/ of "seminar" is not; however, to the extent that lexical constraints enhance phonemic restoration (Samuel, 1987), such effects were not expected to play a role because of the presence of other acoustic cues to /s/ in the speech signal and because the same stimuli were repeated over and over in the course of the experiment. As to the problem of repetition itself, there is no evidence in the literature that repeated presentation reduces phonemic restoration (see Warren & Obusek, 1971; Samuel, in press).

In addition to the two critical words, "seminar" and "happiness," two control words were used which had a /t/ instead of an /s/ in either initial or final position, in the same vocalic context: "telephone" and "cabinet." When a noise target was placed in these words, it occurred during the silent stop closure, so there was no acoustic material for listeners to restore. According to both hypotheses, the noise targets in these stimuli should be perceived veridically. A pilot study revealed a strong bias to perceive target noises as lower in pitch than probe noises. Such a bias should affect all stimuli equally, however. Therefore, the critical comparison in Experiment 1 was between the test sentences containing /s/ and the control sentences containing /t/.

Method

Subjects. Ten volunteers from the Yale community were paid for their participation. They were all native speakers of English and claimed to have normal hearing.

Stimuli. A female speaker produced the four sentences, "Say (seminar, happiness, telephone, cabinet) again," with neutral intonation in a sound-insulated booth. The utterances were recorded using high-quality equipment and were then digitized at a sampling rate of 20 kHz, with low-pass filtering at 9.8 kHz, but *without* high-frequency pre-emphasis. Using a digital waveform editor, the [s] noises of "seminar" and "happiness," 148 and 135 ms in duration, respectively, were excerpted and stored in separate computer files. Their average Fourier spectra were computed and inspected to determine the lower cutoff frequency of the main spectral energy. Each [s] noise was then converted into amplitude-modulated white noise by randomly reversing the polarity of sampling points with a probability of 0.5 (Schroeder, 1968). The resulting noises had exactly the same amplitude envelopes as the original [s] noises, but a flat spectrum. Each of these noises was then passed through a digital low-pass filter whose cutoff frequency was set approximately where the energy of each [s] noise fell off (i.e., around 4 kHz). The attenuation level above the cutoff frequency was varied by trial and error until a noise resulted (N) that, when digitally added to the [s] noise waveform, yielded a noise with an approximately flat spectrum (NS). Before the noises were added together, the amplitude of the filtered noise was adjusted to match that of the natural [s]. A third noise (NSS) was created by adding the [s] waveform to NS with a temporal offset of a few samples, to avoid simply doubling the [s] sampling values in the mixture. The spectra of these noises derived from "seminar" and "happiness" are shown in Figure 1.^{3}

Using the waveform editor, the NS and NSS noises were spliced back into the test sentences where the [s] noise had been. They were also spliced into the silent /t/ closures of "telephone" and "cabinet," which were extended to accommodate the noises. The original durations of the /t/ closures were 67 and 76 ms, respectively. Since each closure coincided with a word boundary, it was reasonable to assume that extending its duration would have a negligible effect on /t/ perception.

Two stimulus sequences were recorded on audio tape. The first sequence served as practice and contained pairs of isolated noises. Pilot tests had suggested that without some training in the noise matching task subjects' performance was very poor. Pairings of N with NSS were omitted on this tape, as they were too easy to discriminate. Thus there were three identical (N-N, NS-NS, NSS-NSS) and four nonidentical (N-NS, NS-N, NS-NSS, NSS-NS) pairs for each of the two sets of noises, 14 pairs in all, which were recorded 12 times in random order, with interstimulus intervals (ISIs) of 750 ms within pairs and 3 s between pairs. The second sequence, the main test, contained sentences paired with noise probes. There were five such combinations for each sentence: NS and NSS as targets paired with either NS or NSS as probes, and NS as target with N as probe. The probe occurred either before or after the sentence, with an ISI of 250 ms. Thus there were 40 possible stimuli altogether, which were recorded four times in random sequence, with ISIs of 3 s.

Procedure. The subjects sat in a quiet room and listened over TDH-39 earphones. Their task was to write down one of three responses for each pair of noises: an up-pointing arrow if the second noise seemed higher in pitch than the first; a down-pointing arrow if it seemed lower; or an equal sign if the two noises seemed identical. The concept of noise pitch was first explained with reference to the phones [s] and [ʃ], [s] having a higher pitch than [ʃ]. It was emphasized that the differences to be detected would be small and that careful listening was necessary. The correct responses to the first 28 practice pairs of isolated noises were already filled in on the answer sheet, and subjects were requested merely to check them off while listening; the subjects then responded to the remaining noise pairs. Before the main test, subjects were told that the task remained the same, but that one of the noises in each pair would coincide with speech that was irrelevant and should be ignored. After a short break, the main test was repeated, so that each subject gave 8 responses to each individual stimulus pair.

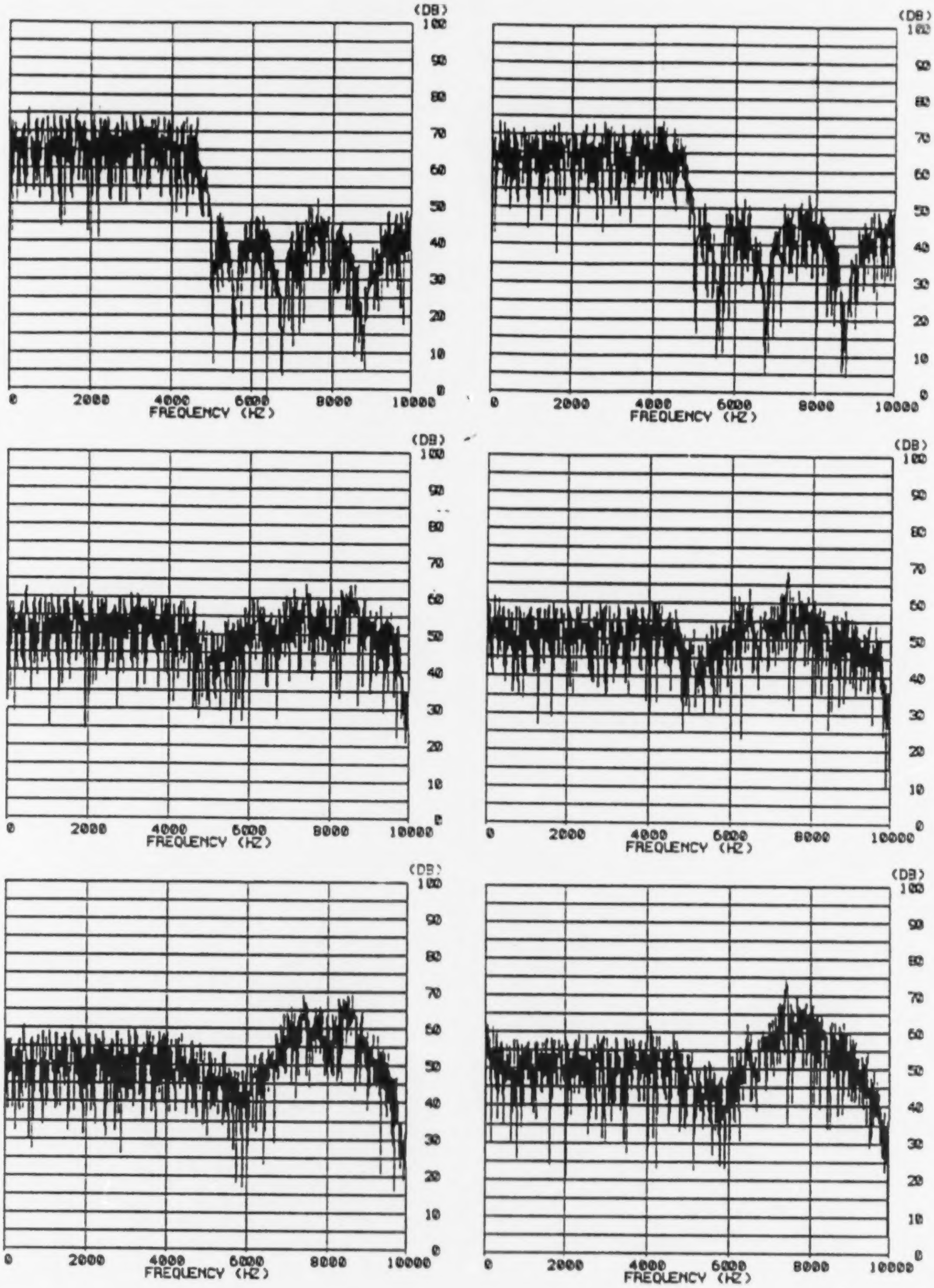


Figure 1. **Fourier spectra (without high-frequency pre-emphasis) of the N (top), NS (middle), and NSS (bottom) noises. The noises for "seminar" are on the left, those for "happiness" are on the right.**

Results and Discussion

The results for the isolated noises are shown in Table 1. The subjects had little difficulty in discriminating the N and NS noises, but they often did not detect the difference between NS and NSS. Accuracy for identical pairs was intermediate. On the whole, these results are satisfactory, but the auditory similarity of NS and NSS must be kept in mind when interpreting the results of the main test.

These results are depicted in Figure 2. Each of the five panels compares the results for control (/t/) and test (/s/) sentences for one of the five probe-target combinations. The four functions in each panel represent the two possible temporal orders of probe and target (P-T or T-P) for each of the two possible locations in the word (initial or final). The quantity plotted on the ordinate is the difference between the percentages of "probe higher than target" and "probe lower than target" judgments. These scores thus range from -100 to 100, with 0 representing perceived identity. ("Equal" judgments were not considered; this way of treating the data is equivalent to assigning weights of -1, 0, and 1 to the three types of responses.) A repeated-measures ANOVA was conducted on the individual subjects' scores, with the N/NS condition omitted to balance the design. There were six crossed within-subject factors in the ANOVA: Condition (/t/ vs. /s/), Probe (NS vs. NSS), Target (NS vs. NSS), Location (initial vs. final), Order (P-T vs. T-P), and Repetitions (first vs. second presentation of the main test).

Table 1. *Average response percentages for isolated noise pairs in Experiment 1 (second noise lower/equal/higher than first noise).*

First noise	Second noise																	
	Initial type									Final type								
	N			NS			NSS			N			NS			NSS		
N	10	82	8	0	2	9	–	–	–	10	88	2	8	3	89	--	–	–
NS	98	0	2	13	76	12	13	40	48	97	3	0	9	88	3	5	44	51
NSS	--	–	–	61	29	10	8	78	14	–	–	–	38	56	6	5	85	10

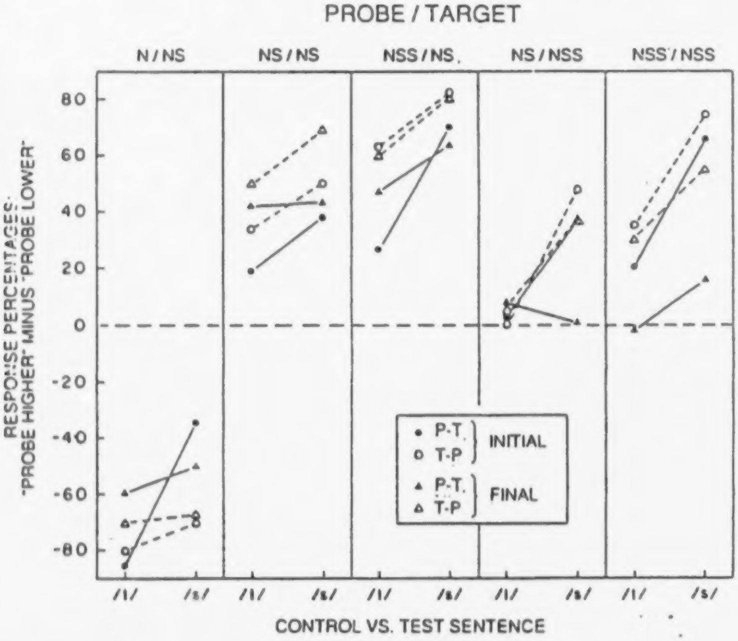


Figure 2. **Results of Experiment 1 (main test). Each panel shows one particular probe-target combination. P-T probe followed by target, T-P target followed by probe; initial word-initial fricative, final word-final fricative. The rising slope of the lines is of main interest.**

The most striking feature of Figure 2, and also the most important result, is that in 19 out of 20 comparisons (individual lines in the figure) the test (/s/) sentence probes received higher scores overall than the control (/t/) sentence probes. The Condition main effect was highly significant, F(1,9) = 23.83, $p < 0.001$. This means that, relative to the control baseline, the targets in the test sentences were perceived as lower in pitch than the targets in the control sentences. This result supports the segregation hypothesis and suggests that the restored [s] noise was perceptually subtracted from the noise in the sentence.

The importance of including the control baseline is underscored by the presence of a strong bias to judge probes as higher in pitch than targets: Seven out of eight scores for physically identical probe-target combinations (NS/NS, NSS/NSS) in control (/t/) sentences were well above zero (see panels 2 and 5 in Figure 2).

Even though the NS and NSS noises were not easy to discriminate in isolation, there were significant main effects of Probe, F(1,9) = 18.35, $p < 0.003$, and of Target, F(1,9) = 9.90, $p < 0.02$, which show that the subjects could distinguish the two noises to some extent even in the main test. Target noises seemed to be more difficult to distinguish than probe noises, due to the surrounding sentence context.[4]

If auditory segregation of [s] from the target noise had been complete, the N probe should have been perceived as matching the NS target in the test sentences; however, the N probe was clearly perceived as still much lower in pitch (panel 1 of Figure 2). The segregation hypothesis also predicted that the NS probe should be perceived as matching the NSS target in the test sentences, but instead the probe was perceived as *higher* in pitch, contrary to the actual physical difference (panel 4 of Figure 2). This was due to the general bias to perceive probe noises as higher in pitch than target noises. Therefore, these results need to be compared with the scores for physically identical probe-target pairs (NS/NS and NSS/NSS) in control sentences; these seem comparable, being subject to the same bias.

Thus, although the predictions of the strong segregation hypothesis are not upheld in one condition (N probe/NS target, NS targets corresponding to the "replaced" items in Samuel's design), they are consistent with the results for another condition (NS probe/NSS target, NSS targets corresponding to Samuel's "added" items). Note, however, that the NS target can also be regarded as resulting from the addition of an extraneous low-pass filtered noise N to [s], whereas the NSS target could have resulted from replacing the [s] noise with a spectrally similar noise NSS. Ultimately, therefore, the distinction between "added" and "replaced" becomes meaningless, and what takes its place is the continuous dimension of *spectral similarity* between the target noise and the original [s] noise (or, rather, the [s] noise that a listener might expect to occur, since listeners never actually heard the original [s] noise in this experiment, though they heard an /s/ in the word "say" at the beginning of each sentence). It seems, then, that [s] restoration is more complete when the target noise is spectrally similar to an [s] noise (NSS) than when it is not (NS). Listeners seem to perceptually subtract more from a NSS target than from a NS target. In support of this suggestion, the author observed informally as a listener that, even though /s/ restoration clearly occurred in both conditions, the /s/ in the NS target condition seemed less natural than that in the NSS target condition; it had a "lisping" quality. Phone restoration thus seemed to be incomplete in the NS target condition, even though the original [s] noise was physically present in the target noise. This observation will be discussed further in the General Discussion.

EXPERIMENT 2

The results of Experiment 1 were highly suggestive of auditory segregation, but there were some irregularities in the results, and a replication seemed desirable. Experiment 2 provided a partial replication using probes of a finer grain. Since so few data points fell close to zero in Experiment 1 (see Figure 2), it was of interest to determine more precisely what kind of probe was judged to match the target, particularly for the NS target condition. This was achieved by constructing a continuum of five probe noises between N and NS, and using two of them as targets. Because of the increase in probe-target combinations, only the initial-location stimuli ("seminar" and "telephone") were used, which had yielded more consistent data in Experiment 1.

One additional change in Experiment 2 concerned the instructions for the repetition of the main test. Whereas, in the first presentation, subjects were asked to ignore the speech, as in Experiment 1, during the second presentation they were required to identify the critical word in the sentence by writing down its initial letter after making their judgment about the noise pair. This additional requirement was intended to increase subjects' attention to the speech. The question was whether this would affect the extent of auditory segregation. For example, it could be that phone restoration in Experiment 1 was incomplete because attention was directed away from the speech, toward the extraneous noise, even though this seems unlikely.

Method

Subjects. Ten paid subjects from the same general population participated.

Stimuli. The stimuli, derived from "seminar" and "telephone," were the same as in Experiment 1, with the exception of the following changes in probe and target noises: The N and NS noises were renamed N1 and N5, respectively, and three additional noises with intermediate spectra (N2, N3, N4) were created by adding the digitized waveforms (matched for overall amplitude) of N (the low-pass filtered noise) and the original [s] noise with the following weights: 0.875/0.125, 0.75/0.25, and 0.625/0.375 (corresponding approximately to N/[s] ratios of 17, 9.5, and 4.5 dB). N5, of course, represented the addition of N and [s] with equal weights of 0.5 (N/[s] ratio of 0 dB). These noises thus constituted a series of increasing pitch. All five noises were used as probes; however, only N3 and N5 were used as targets. The asymmetry between probe and target noises (N1 was not used as a target), anticipated (but possibly also encouraged) the bias to judge probes as higher in pitch than targets.

As in Experiment 1, two test sequences were recorded. The first, for practice, contained pairs of isolated noises with ISIs of 750 ms within pairs and 3 s between trials. N3 and N5 occurred in all possible pairings with each other and with all other noises: seven nonidentical pairs (in two possible orders) and two identical pairs (repeated once), a total of 18 stimulus pairs. Six different randomizations of these 18 pairs were recorded. The second test sequence, the main test, contained five randomizations of 40 noise pairs. One noise in each pair (the target) was buried in one of the two speech contexts; the other noise (the probe) preceded or followed the speech with an ISI of 250 ms. The intertrial interval was 3 s. The 40 pairs resulted from five probes paired with two targets in two speech contexts in two temporal orders. The tape was played twice.

Procedure. The procedure and instructions were exactly the same as in Experiment 1 except that, for the second presentation of the main test, subjects were asked to write next to their judgment of the noises the letter S or the letter T, depending on whether the speaker on the tape had said "seminar" or "telephone."

Results and Discussion

The results for the isolated noise pairs are shown in Table 2. It can be seen that the subjects, while far from perfect, were clearly sensitive to the spectral differences among the noises, and their accuracy was quite satisfactory for the purposes of this experiment.

The results of the main test, pooled over the two repetitions, are shown in Figure 3. The response measure is the same as in Figure 2; the higher the score, the higher in pitch the probe was judged relative to the target, with zero representing perceived equality. On the abscissa we have now a continuum of probe noises, with separate panels for preceding and following probes. It is evident that subjects' responses increased as an orderly function of probe spectrum, as in the pretest [$F(4,36) = 78.11$, $p < .0001$, for the Probe main effect]. Probe discrimination seemed slightly better when the probe followed than when it preceded the speech, as indicated by the steeper functions in the right-hand panel of Figure 3 [$F(4,36) = 5.74$, $p < .002$, for the Probe by Order interaction]. Clearly, the two targets were also discriminated, with higher ratings for probes paired with N3 than for probes paired with N5 [$F(1,9) = 34.31$, $p < 0003$, for the Target main effect]. This difference was somewhat more pronounced when the probe preceded the target, i.e., when the target did not have to be held in memory [$F(1,9) = 13.48$, $p < .006$, for the Target by Order interaction]. Most importantly, however, there was a consistent difference between the /s/ and /t/ conditions: For all probe-target combinations, the probe was judged higher (hence, the target was judged lower) in pitch when [s] was to be restored, in agreement with the segregation hypothesis [$F(1,9) = 83.27$, $p < .0001$, for the Condition main effect]. The difference was somewhat larger when the probe preceded the target [$F(1,9) = 12.49$, $p < .007$, for the Order by Condition interaction]. There were also significant Target by Probe [$F(4,36) = 3.76$, $p < .02$] and Condition by Target by Probe interactions [$F(4,36) = 4.27$, $p < .007$], reflecting differences in the shapes of the response curves due to floor and ceiling effects. (No data transformation was applied.) The analysis of variance further revealed a Repetitions by Order interaction [$F(1,9) = 7.64$, $p < .03$], due to an increase in the Order effect in the second half of the main test, and a marginally significant four-way interaction. On the whole, however, instructions to pay attention to the speech in the repetition of the main test had little effect (nor did practice or fatigue, for that matter).

Table 2. *Average response percentages for isolated noise pairs in Experiment 2 (second noise lower/equal/higher than first noise).*

First noise	**Second noise**														
	N1			**N2**			**N3**			**N4**			**N5**		
N1	–	–	–	–	–	–	2	10	88	–	–	–	0	0	100
N2	–	–	–	–	–	–	10	36	54	–	–	–	2	0	98
N3	98	0	2	78	16	6	12	70	18	10	46	44	0	24	76
N4	–	–	–	–	–	–	32	62	6	–	–	–	6	30	64
N5	98	0	2	98	2	0	80	18	2	60	14	26	4	88	8

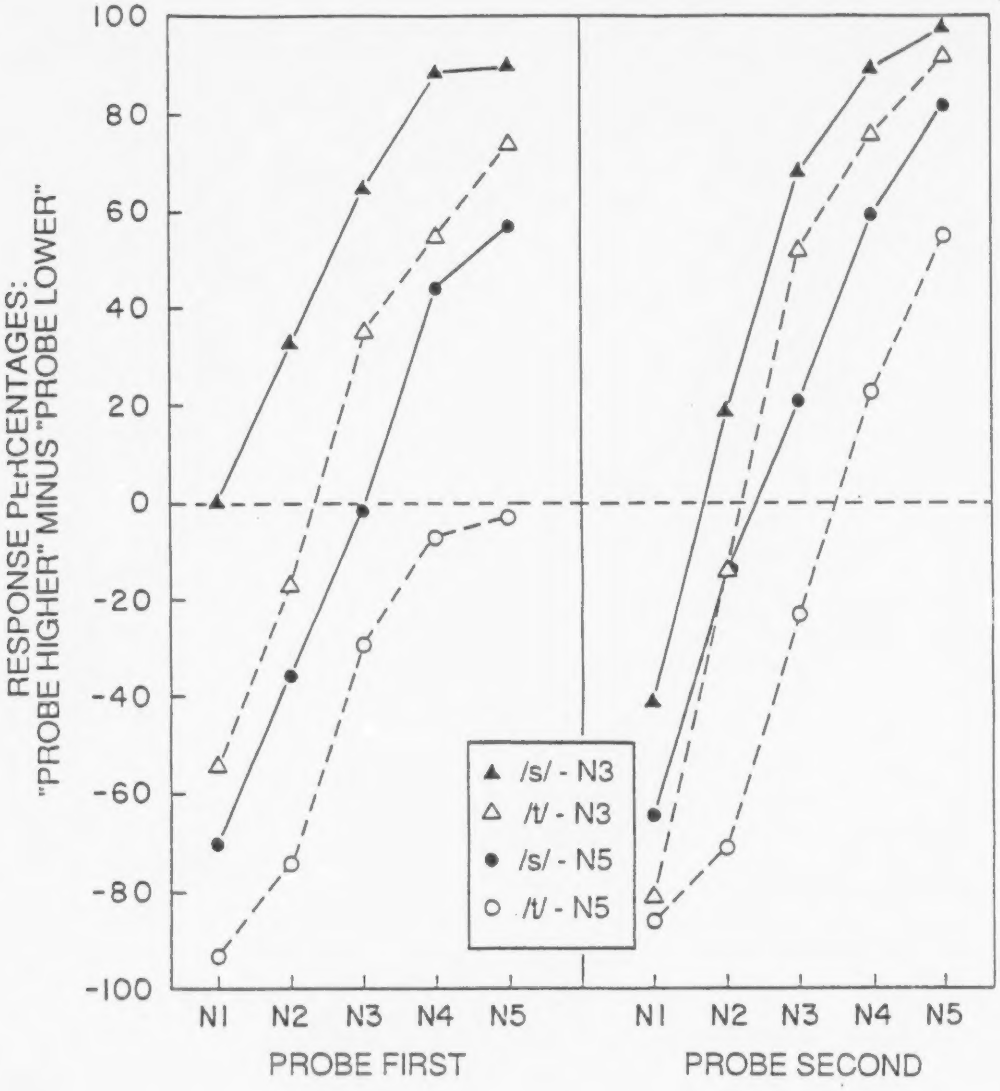


Figure 3. **Results of Experiment 2 (main test). N3 and N5 (legend) are the target noises; /s/ and /t/ are the phonemes to be restored.**

Thus the results of this experiment again lend support to the segregation hypothesis, even for a target noise (N3) that was rather dissimilar to [s]. As in Experiment 1, however, the apparent phone restoration was far from complete. If it had been, the N5 target should have been judged as matching the N1 probe; in fact, it was judged closer to N3 or perhaps N2 (filled circles in Figure 3). A large part of that shift, moreover, must be attributed to the general bias to perceive probes as higher in pitch, which was again present: In three out of four pairings of probes with identical targets replacing /t/, the probe was judged considerably higher in pitch than the target (N3/N3 and N5/N5 probe/target pairs, open symbols in Figure 3). If this bias is taken into account, the N5 target seemed most similar to the N3 or N4 probe. Thus the [s] segregation effect, though consistently present, was relatively small. Clearly, listeners did not extract a full [s] from the target noise.

EXPERIMENT 3

The initial two experiments have provided evidence of partial segregation of an [s] that has been replaced with a different noise. The question must be raised now, however, whether this apparent segregation was induced by top-down expectations (as in the standard phoneme restoration paradigm) or by residual cues to the "missing" phone in the acoustic signal. Even though the formant transitions in adjacent vocalic segments were not highly distinctive between /s/ and /t/, and cues to fricative manner were weak, the presence of the /t/ release burst plus aspiration provided salient information for the presence of a /t/ in the control sentences. The difference between the /s/ and /t/ conditions, on which the conclusions of Experiments 1 and 2 rest, could thus be due to an inhibition of [s] restoration by local acoustic context (rather than by lexical constraints) in the control condition or, what is worse, it could even reflect some direct effect of acoustic environment on the perceived pitch of the target noise.

If phone restoration were facilitated or inhibited by acoustic environment, this would be an interesting finding in itself. However, a vital concern of current models of word recognition (e.g., McClelland & Elman, 1986; Marslen-Wilson, 1987) is whether lexical top-down effects can penetrate to relatively early levels of speech processing. Elman and McClelland (1988) have argued that they can, after showing that a lexically disambiguated (they call it "restored") fricative can influence the perception of a following stop consonant. (See also Mann & Repp, 1981, who reported essentially the same finding for spontaneous fricative disambiguation.) *Phoneme* restoration was lexically driven in the classic experiments in which all acoustic cues were removed. Experiment 3 examined whether *phone* restoration and auditory segregation would be observed when differences in bottom-up contextual cues are eliminated. This was accomplished by making the acoustic environment of the target noise in "telephone" identical to that in "seminar," favoring either /s/ or /t/ in either context. If bottom-up factors were solely responsible for the segregation effect, then it should disappear under these conditions.

A second question addressed in Experiment 3 concerned the strong bias to perceive target noises as lower in pitch than probe noises. One possible origin of this tendency is that the /t/ environment, because of its relative similarity to that of /s/ and because of the artificial lengthening of the stop closure interval, actually encouraged perception of /s/ (only less so than the /s/ environment), even though the /t/ was aspirated and a nonword was the result ("selephone" or "stelephone"). In Experiment 3, therefore, yet another control condition was employed, in which the target noise not only replaced a segment (a /p/ closure) whose acoustical context was dissimilar from that of either /s/ or /t/, but also was followed by a segment (/r/) that phonotactically barred the occurrence of an immediately preceding /s/. If the bias were absent in that condition, then it would have to be attributed to /s/ restoration in the /t/ condition.

Method

Subjects. Ten new subjects from the same general pool participated.

Materials. A new set of six utterances was recorded by a different female speaker. The carrier phrase was once more "Say ... again," and the embedded items were "seminar," "teminar," "selephone," "telephone," "president," and "tresident." The first four items thus varied lexical status (word or nonword) and local acoustic information (/s/ or /t/ cues) independently, whereas the last two items introduced a phonotactic bias against /s/ while also varying local information and lexical status.

Target and probe noises (N1-N5) were constructed exactly as in Experiment 2, using as the starting point a 128-ms [s]-noise excerpted from another utterance containing "selephone." Again, only N3 and N5 were used as targets, whereas all five noises were used as probes. The pretest contained 12 random sequences of 18 pairs of isolated noises (twice as many as in Experiment 2, to provide additional practice). The main test contained 5 blocks of 60 target-probe pairs (6 items × 2 targets × 5 probes). There were two important differences with respect to Experiment 2: First, to reduce the number of stimulus pairs, the probe noise always *followed* the target. Second, in the items originally containing a stop consonant, the target noise replaced the closure *and the release burst*; the aspiration following the release burst was left intact but provided a much weaker cue to stop consonant manner.

Procedure. As in Experiment 2, the subjects listened first to the pretest, with correct answers provided for the first two blocks. The main test was presented twice. During the second presentation, the subjects were required to write down, in addition to their judgment of the noises, the initial consonant of the item containing the target. They were told that the speaker sometimes mispronounced the words. The two possible choices (S/T for "-eminar" and "-elephone," and P/T for "-resident") were listed on top of the answer sheet.

Results and Discussion

The pretest results are shown in Table 3, and they resemble those obtained in Experiment 2. Clearly, the subjects were able to discriminate the noises quite well in isolation.

The results of the main test are presented in Table 4. The response functions (rows in the table) were steep; clearly, the probe noises were discriminated well by the subjects [$F(4,36) = 165.96$, $p < .0001$, for the Probe main effect]. There was also a reliable difference between the two target conditions, though it was small in size [$F(1,9) = 24.31$, $p < .0009$]. However, there were no other interesting differences in the data. Several other significant effects in the ANOVA were all due to an unexplained anomaly: a reversal of the target effect for the "telephone" items (see Table 4, N3 probe).

Table 3. *Average response percentages for isolated noise pairs in Experiment 3 (second noise lower/equal/higher than first noise).*

First noise	Second noise														
	N1			N2			N3			N4			N5		
N1	--	--	--	--	--	--	4	16	80	--	--	--	2	0	98
N2	--	--	--	--	--	--	8	52	40	--	--	--	2	2	96
N3	98	2	0	88	8	4	16	77	7	2	46	52	4	6	90
N4	--	--	--	--	--	--	70	24	6	--	--	--	16	42	42
N5	100	0	0	96	2	2	82	12	6	16	64	20	2	80	18

Table 4. *Average scores (percent judgments of "probe higher" minus percent judgments of "probe lower") in the main test of Experiment 3. Also, in the last column, the average percentages of consonant identification responses.*

Item	Target	Probe					Percent responses		
		N1	N2	N3	N4	N5	S	T	P
seminar	N3	- 85	- 47	22	68	84	77	23	--
	N5	- 84	- 57	-1	52	87	97	3	--
teminar	N3	- 84	- 42	14	69	89	36	64	--
	N5	- 93	- 65	- 23	57	81	60	40	--
selephone	N3	- 78	- 39	11	66	91	89	11	--
	N5	- 88	- 66	- 11	61	89	96	4	--
telephone	N3	- 69	- 46	1	75	86	12	88	--
	N5	- 85	- 50	17	74	85	19	81	--
president	N3	- 86	- 42	15	69	90	--	14	86
	N5	- 84	- 64	- 8	45	74	--	21	79
tresident	N3	- 69	- 41	6	56	91	--	42	58
	N5	- 86	- 57	- 12	34	83	--	54	46

The prediction of the segregation hypothesis was that scores would be higher for /s/ items than for /t/ items, and in particular higher for "seminar" than for "telephone," as they were in Experiment 2. However, these differences were small and nonsignificant, even slightly in the opposite direction. Thus the results did not replicate the principal finding of Experiments 1 and 2, viz., that [s] restoration affected perception of the extraneous noise, nor did they demonstrate any systematic effect of local acoustic environment. They did show, however, the by now familiar bias to perceive probes as higher in pitch than targets, though only for N5 targets, which were perceived as similar to N3 probes. N3 targets, on the other hand, were also perceived as most similar to N3 probes.

What phonemes did the subjects actually restore? In Experiments 1 and 2, correct phoneme restoration was taken for granted because lexicality and acoustic cues coincided. In the present experiment, these two factors were dissociated. Were subjects guided by lexicality or by acoustic cues or by both? A tally of subjects' initial consonant responses in the second presentation of the main test provided an answer. The average percentages of consonant responses are shown in the last three columns of Table 4. Each subject gave 25 responses to each item. There was considerable variability in subjects' identification of the items. It should be noted first, however, that the two items that had also been used in the previous experiments, "seminar" and "telephone," were generally identified correctly (i.e., as intended by the speaker of the materials). Although a substantial number of "errors" did occur, they are not sufficient to explain the total absence of a segregation effect in the present experiment. It can further be seen in Table 4 that the items "selephone" and "president" were generally identified correctly. Not so for "teminar" and "tresident," however, which tended to be identified as "seminar" and "president," respectively, about half of the time. In addition to these differences among items, which apparently reflect both acoustic and lexical factors, an influence of the target noise on consonant identification can be seen: In the first four items, where S and T were the response choices, N3 (the lower-pitched noise) favored T, while N5 (the higher-pitched noise) favored S. In the last two items, where P and T were the choices, N3 favored P and N5 favored T. Since the noisy acoustic correlates of /p/, /t/, and /s/ form a series of increasing pitch, the influence of the target noise spectrum on phoneme restoration may represent an instance of "bottom-up confirmation" (Samuel, 1981b).

The effects just described were statistically reliable. A 2 × 2 × 2 ANOVA was conducted on the S response percentages to the the first four items ("s/teminar," "s/telephone"), with the factors Context ("-eminar" vs. "-elephone"), Cues (/s/ vs. /t/), and Target (N3 vs. N5). The main effects of Cues and of Target were highly significant [$F(1,9) = 40.54, p < .0002$; $F(1,9) = 61.12, p < .0001$]. The Context by Cues interaction was likewise significant [$F(1,9) = 31.88, p < .0004$] and probably reflects an effect of lexicality (words vs. nonwords). In addition, there was a Context main effect [$F(1,9) = 12.39, p < .007$], due to more S responses to "s/teminar" than to "s/telephone," and a Context by Target interaction [$F(1,9) = 6.08, p < .04$], due to a larger target effect for "s/teminar" than for "s/telephone." A separate 2 × 2 ANOVA on the "s/tresident" items revealed significant Cues and Target main effects [$F(1,9) = 9.99, p < .02$; $F(1,9) = 23.36, p < .001$]. Lexicality was confounded with Cues here.

In summary, these data show that the identity of the restored phoneme is influenced by both acoustic and lexical factors. The acoustic determinants in particular are very clear: Both the residual cues in the signal (despite elimination of the stop release bursts) and the spectral properties of the target noise had effects. The effect of lexicality seems less systematic because of the accurate identification of "selephone."[5] In the other two comparisons, the real words ("seminar," "president") were identified more accurately than the nonwords ("teminar," "tresident"), and so was the real word "telephone" which shared the initial consonant with these nonwords. Since all items became thoroughly familiar to the subjects in the course of the experiment, lexical biases were presumably attenuated compared to a more natural situation.

EXPERIMENT 4

Even though the subjects generally identified the critical test items as intended, the segregation hypothesis was not supported in Experiment 3. The subjects also had some difficulty distinguishing the target noises in the main test. One disturbing possibility is that the subjects may have adopted a strategy of essentially ignoring the target noise and making their noise comparison judgments on the basis of the probe noise alone, responding "higher" when its pitch was high and "lower" when its pitch was low. This strategy may have been encouraged by the fact that the probe noise always followed the sentence, so that the target had to be held in auditory memory for comparison, whereas the probe noise was freshly perceived. In Experiment 4, therefore, the design of the earlier experiments, with probes both preceding and following the targets, was restored. To reduce the number of stimuli, the N1 and N5 probes, which provided little information, were dropped. The replication experiment was also desirable because of the anomalous data for the "telephone" stimuli in Experiment 3, and because of the restrictive two-alternative choice task in consonant identification. In Experiment 4, the subjects were free to write down any initial consonant they heard during the second half of the main test.

Method

Subjects. Eleven new subjects were recruited. Three subjects' data were dropped, however, because of poor performance in the pretest or the main test, which left only 8 subjects.

Materials. The stimuli were the same as in Experiment 3. A new main test was constructed in which the probe noise either preceded or followed the sentence. To compensate for the doubling in the number of trials, only N2, N3, and N4 were used as probes. This resulted in a test comprising 5 blocks of 72 target-probe pairs (6 items × 2 targets × 3 probes × 2 orders).

Procedure. The procedure was the same as in Experiment 3, except that, in the second presentation of the main test, the choices for consonant identification were not listed, and subjects were free to write down whatever consonant they heard.

Results and Discussion

The pretest of isolated noises was the same as in Experiment 3, and the results for it are shown in Table 5. Performance was comparable to that in Experiment 3.

The results of the main test are shown in Table 6. It is evident that the subjects performed very poorly when the probe noises preceded the target, in contrast to Experiment 2, for reasons that are not quite clear. A separate ANOVA on this half of the data showed neither a significant main effect of Probe [F(2,14) = 1.13, $p < .35$] nor one of Target [F(1,7) = 1.32, $p < 28$]. The only significant effects were a Target by Repetitions interaction [F(1,7) = 8.32, $p < .03$], suggesting that the subjects began to discriminate the target noises in the second half of the test, and a marginal higher-level interaction also involving Repetitions. Thus, this part of the data was rather useless.

Fortunately, the results for following probes look better, at least as far as subjects' sensitivity to the probe noises is concerned. The ANOVA on this part of the data (in which the 6 items were treated for convenience as representing a 3 × 2 Context by Cues design, even though the cues in the "p/tresident" pair were different) showed a highly significant main effect of Probe [F(2,14) = 18.27, $p < .0002$], and the subjects' response functions (truncated here) were comparable to those obtained in Experiments 2 and 3. However, there were no other significant effects. In particular, the Target main effect was nonsignificant [F(1,7) = 1.36, $p < .28$], indicating that the subjects were unable to discriminate the target noises, and the Cues and Context main effects, as well as their interaction, were nonsignificant, indicating neither acoustic nor lexical influences on perception of the target noise as such.

Table 5. *Average response percentages for isolated noise pairs in Experiment 4 (second noise lower/equal/higher than first noise).*

First noise	Second noise														
	N1			N2			N3			N4			N5		
N1	--	--	--	--	--	--	10	15	75	--	--	--	3	0	97
N2	--	--	--	--	--	--	5	48	47	--	--	--	0	5	95
N3	100	0	0	92	3	5	14	67	19	5	33	62	3	5	92
N4	--	--	--	--	--	--	59	28	13	--	--	--	3	35	62
N5	97	0	3	95	0	5	67	23	10	25	67	8	4	61	35

Table 6. *Average scores (percent judgments of "probe higher" minus percent judgments of "probe lower") in the main test of Experiment 4.*

Item	Target	Probe first			Probe second		
		N2	N3	N4	N2	N3	N4
seminar	N3	-21	1	6	- 45	-1	53
	N5	3	5	13	- 45	-8	38
teminar	N3	-16	-15	5	- 21	-1	38
	N5	-13	3	14	- 53	8	46
selephone	N3	-15	-14	-6	- 33	-11	49
	N5	1	-6	26	- 39	-6	48
telephone	N3	-35	-4	9	- 44	24	44
	N5	-11	6	3	- 51	18	34
president	N3	-8	-1	8	- 40	6	53
	N5	-6	5	16	49	11	51
tresident	N3	-10	8	6	- 24	-1	59
	N5	-8	14	8	- 43	11	31

In a further attempt to detect any systematic tendencies in the data, an ANOVA was performed on the combined data of Experiments 3 (omitting data for the N1 and N5 probes) and 4 (data for following probes only). Because of the increased statistical power, a number of effects were significant in this analysis. Thus there were the expected main effects of Probe [$F(2,34) = 177.25$, $p < .0001$] and of Target [$F(1,17) = 18.08$, $p < .0006$], as well as a number of significant interactions. The only interaction of potential interest, however, between Context, Cues, and Probe [$F(4,34) = 4.50$, $p < .003$], was due to a complex and essentially uninterpretable pattern of results. The main prediction of the segregation hypothesis, that scores should be higher in items with /s/ cues than in items with /t/ cues, was not supported, and neither the main effects of Context and Cues nor their interaction were significant.

Table 7 shows how the subjects in Experiment 4 identified the restored initial consonants, without any restraint on their choices and without any information on how the words might have been "mispronounced" by the speaker. The data must be viewed with some caution because listeners differed substantially in their reactions to these stimuli. (One subject, for example, identified all items as intended by the speaker.) What the results show, however, is that responses other than the expected ones (S, T, P) were quite frequent, particularly F (for "-resident") and TH (for "-elephone"); H responses to stimuli originally starting with /t/ may have reflected listeners' perception of the aspiration noise remaining in the signal. There are some striking differences from the forced-choice results of Experiment 3, in that the initial consonants of "-eminar" and "-resident" were *never* identified as T, and even "telephone" received fewer T responses than previously. Clearly, restoration of stop consonants was not encouraged by the target noise, and fricative responses predominated. There was no clear effect of residual acoustic cues here, except in "-elephone," where T responses were more frequent to "telephone" than to "selephone"; this could also have been a lexical bias, however. S responses were more frequent to "seminar" than to "selephone," whereas T responses were more frequent to "telephone" than to "teminar," which may be an effect of lexicality. There was also a small but consistent effect of target noise; as in Experiment 3, the higher-pitched noise (N5) encouraged /s/ restoration. Because of the high intersubject variability, however, no statistical analysis was conducted on these data.

As to the two items shared with Experiment 2, these results show very clearly that /s/ restoration did occur in "seminar," whereas listeners were less certain about the initial consonant of "telephone" but reported /s/ only one third of the time. Nevertheless, as in Experiment 3, there was no significant difference in subjects' perception of the extraneous noise in these items. Taken together, the data of Experiments 3 and 4 raise the possibility that the results of Experiments 1 and 2, which had been interpreted as supporting the segregation hypothesis, were in fact not caused by [s] segregation. The "telephone" stimuli of Experiments 1 and 2 differed from those of Experiment 3 in that they retained the /t/ release burst in the acoustic signal. This high-frequency burst may have affected subjects' perception of the target noise in an assimilative fashion, making the noise seem higher in pitch, or it may itself have been treated as part of the noise to be judged.

Table 7. *Average percentages of consonant identification (open choice) in Experiment 4.*

Item	Target	Response						
		S(C)	T	P	F	TH	H	Other[a]
seminar	N3	91	0	0	8	0	0	0
	N5	98	0	0	1	0	0	0
teminar	N3	85	0	0	7	0	6	1
	N5	95	0	0	3	0	2	0
selephone	N3	61	17	0	7	16	0	0
	N5	68	15	0	7	10	0	0
telephone	N3	30	40	0	9	14	5	2
	N5	35	39	0	8	15	3	1
president	N3	0	0	43	57	0	0	0
	N5	1	0	41	58	0	0	0
tresident	N3	0	0	40	59	1	1	0
	N5	0	0	40	59	0	0	0

[a] A few each of ST, STH, SH, CH. Rows may not add up to 100 percent because of rounding error.

The results of Experiments 3 and 4 offer no support for the segregation hypothesis; by suggesting that phone restoration (to the extent that it occurs at all) leaves perception of the extraneous noise unaffected, they seem to favor the top-down completion hypothesis.

In addition, Experiments 3 and 4 show that the general bias to perceive the target noise as lower in pitch than the probe noise was not due to [s] restoration in stimuli meant to induce restoration of some other phone. The bias was equally present in "-resident" stimuli, which never led to [s] restoration. Although /f/ percepts were common in these stimuli, [f] frication is lower in both pitch and intensity than [s], which should have been reflected in perception of the target noise if segregation had been operative. Instead, the results suggest that the bias represents a general effect of acoustic context on timbre perception: The perception of high-frequency spectral components may be selectively impaired when a noise is embedded in other sounds.

EXPERIMENT 5

In view of the negative results of Experiments 3 and 4, it seemed desirable to obtain further data, using different stimuli. One thing all experiments so far had in common was the repetition of a limited set of speech stimuli to obtain stable results for each. Although there was reason to believe that extensive exposure and familiarity with the stimuli would not reduce the strength of phone (or phoneme) restoration, the possibility cannot be dismissed that the desired phenomenon was perhaps not elicited in its full strength. In the typical phoneme restoration experiment (as exemplified by Samuel's work), many different lexical items are included. Experiment 5 followed this design.

The crucial comparison in Experiment 5 was between stimuli favoring restoration of /s/ and stimuli favoring restoration of /ʃ/. The corresponding fricative noises are known to differ in pitch, with [ʃ] being lower than [s] on this dimension. Consequently, the segregation hypothesis would be supported if target noises were judged to be lower in pitch when [s] is restored than when [ʃ] is restored.

To induce restoration of these fricative phones, lexical, acoustic, and phonotactic factors were manipulated in the stimuli. Half the stimuli contained /s/ cues and half contained /ʃ/ cues in the phonetic context. Within each of these sets, some stimuli were lexically ambiguous (i.e., they formed a word if the initial consonant was either /s/ or /ʃ/), some were lexically unique (i.e., only one of the two fricatives made a word), and some of these had an additional phonotactic restriction against nonword percepts (the second consonant was /l/ or /r/; word-initial /sl/ and /ʃr/ are common in English, but /sr/ and /ʃl/ are not). If segregation were operative, it should be evident most strongly in this last group of stimuli, which most strongly induce restoration of the intended fricative.

Method

Subjects. Ten new subjects from the same general population were tested.

Stimuli. Eighty familiar, monosyllabic words were selected from a dictionary; they are shown in Table 8. Half of them began with /ʃ/ and half with /s/. As can be seen in the table, half of the words formed minimal pairs distinguished only by the initial consonant (columns 1 and 2). The remaining words had no lexical counterpart beginning with the other fricative (columns 3 and 4). In half of these words, the fricative was followed by a vowel, so that the nonword resulting from substitution of the other fricative was phonotactically regular (e.g., "sade," "shafe"). In the other half, the following consonant was a liquid, /r/ after /ʃ/ and /l/ after /s/, so that nonwords resulting from fricative substitution would be phonotactically irregular (e.g., "srewd," "shlash"). This was expected to introduce a strong bias against perception of such nonwords in the experiment.

The 80 words were spoken by a male speaker in the carrier phrase "Say ... again," in random order and with neutral intonation. The recordings were digitized at 20 kHz without pre-emphasis, lowpass filtered at 9.8 kHz, and edited to remove the fricative noise at the onset of each test word. The durations of these original noises ranged from 120 to 226 ms. One of the noises (from "sack," 178 ms long) served as the basis for generating the target and probe noises, which were slightly different from those used in the previous experiments. The noise was first converted into signal-correlated noise. This envelope-matched broadband noise was used as the target noise in all stimuli, in the hope that the diversity of the materials would prevent the subjects from becoming conscious of the fact that the target noise remained constant. Four additional noises were generated by low-pass filtering the target noise with a cutoff frequency of 5 kHz, using high-frequency stop-band attenuation levels of 12, 9, 6, and 3 dB. These probe noises, together with the original broadband noise which was also used as a probe, formed a continuum of increasing pitch. They will again be referred to by the labels N1 to N5, with N1 standing for the probe noise with the lowest pitch, and N5 standing for the original broadband noise that was also used as the target noise. In this experiment, therefore, there were no probe noises that were actually higher in pitch than the target.

Two test tapes were generated. The pretest, as usual, contained pairs of isolated noises, arranged in 6 blocks of 18, with ISIs of 750 ms within pairs, of 3 sec between pairs, and of 6 sec between blocks. N3 and N5 were paired with each other and with all other noises, just as in Experiments 3 and 4. In the main test, each of the 80 words appeared 5 times, once with each probe noise. Half the words (the odd items in Table 8) appeared with the probe noise preceding the sentence, and the others (the even items) appeared with the probe noise following. They were arranged into 5 blocks of 80 according to a Latin square design, such that each word appeared exactly once in each block, and each probe condition occurred 8 times in each block. The sequence was randomized within blocks, with ISIs of 500 ms within pairs (i.e., between the probe noise and the sentence onset or offset), 3 sec between pairs, and longer intervals between blocks.

Table 8. *Stimulus words used in Experiment 5.*

shack	sack	shade	safe
shag	sag	shaft	salt
shame	same	shape	sand
shave	save	share	seal
sheep	seep	sharp	send
sheet	seat	shawl	silk
sheik	seek	shirt	size
shelf	self	shout	soft
shell	sell	shove	soil
shift	sift	shut	surf
shin	sin	shred	slam
shine	sign	shrewd	slash
ship	sip	shriek	slave
shock	sock	shrill	sleep
shoe	sue	shrimp	slide
shoot	suit	shrine	sling
shore	sore	shrink	slip
short	sort	shroud	slow
show	sew	shrub	slot
shy	sigh	shrug	slur

Procedure. For the pretest, the responses for the first block (18 trials) were filled in on the answer sheet, as in previous experiments. The five blocks of the main test were presented in different orders for different subjects by varying the starting block. After the main test, the first block was presented once again, and the subjects' task was to identify the initial consonant of each word as either /s/ or /ʃ/. Although /s/ and /ʃ/ were suggested as responses, subjects were encouraged to write down any other consonants they might hear. Also, it was emphasized that subjects should write down what they heard, even if that resulted in a nonword.

Results

The results of the pretest are shown in Table 9. The noises were easier to discriminate than those in the previous experiments, or else this group of subjects was just more accurate.

The results of the main test are shown in Table 10. The 80 words were divided into 8 groups of 10, according to whether they originally began with /s/ or /ʃ/ (the Cues factor), whether they were lexically ambiguous or unambiguous (Ambiguity), and whether the probe noise came first or second (Order). The division of nonambiguous words into phonotactically constrained and unconstrained ones is not reflected in the table. There were two highly significant effects: the main effect of Probe [F(4,36) = 74.16, *p* < .0001] and the interaction of Probe with Order [F(4,36) = 22.03, *p* < .0001]. The subjects were quite successful in discriminating the probe noises regardless of whether they came first or second, but they were more accurate when they followed the sentence, as in Experiment 2. The only other effect that reached significance was the Ambiguity by Probe interaction [F(4,36) = 3.11, *p* < .03], apparently due to a tendency to perceive the target noise as relatively higher in ambiguous stimuli, but only when the probe noise was N4. This was the probe noise that, on the whole, was perceived as most similar to the target noise (N5), which again replicates the tendency to perceive target noises as lower in pitch than probe noises. (The present design certainly encouraged such a bias.) The Cues main effect was nonsignificant [F(1,9) = 0.70], though the crucial Cues by Ambiguity interaction was close to significance [F(1,9) + 4.84, *p* < .06]. This interaction was due to one category of words, ambiguous /s/ words, differing from the rest: In these stimuli, the target noise was judged to be relatively *higher* than in the other stimuli. This is contrary to expectations, of course: The segregation hypothesis predicted that target noises would be perceived as lower in /s/ words, particularly in unambiguous /s/ words. Although target noises were perceived as somewhat lower in unambiguous /s/ words than in ambiguous /s/ words, there was no tendency to perceive them as lower in /s/ words than in /ʃ/ words. Thus, the results of this experiment, like those of Experiments 3 and 4, offer no support for the segregation hypothesis.

Table 9. *Average response percentages for isolated noise pairs in Experiment 5 (second noise lower/equal/higher than first noise).*

First noise	Second noise														
	N1			N2			N3			N4			N5		
N1	--	--	--	--	--	--	8	37	55	--	--	--	0	0	100
N2	--	--	--	--	--	--	2	69	29	--	--	--	0	0	100
N3	94	6	0	71	27	2	13	78	9	2	63	35	0	0	100
N4	--	--	--	--	--	--	65	3	4	--	--	--	2	10	88
N5	100	0	0	100	0	0	94	4	2	65	35	0	4	91	5

Table 10. *Average scores (percent judgments of "probe higher" minus percent judgments of "probe lower") in the main test of Experiment 5. The target noise was N5.*

Words	Probe first				
	N1	N2	N3	N4	N5
ambiguous /ʃ/	-60	-47	-22	-14	36
ambiguous /s/	-58	-52	-38	-23	40
unambiguous /ʃ/	-50	-48	-27	-3	47
unambiguous /s/	-64	-52	-32	-5	37
	Probe second				
	N1	N2	N3	N4	N5
ambiguous /ʃ/	-89	-69	-24	24	89
ambiguous /s/	-95	-76	-27	4	76
unambiguous /ʃ/	-90	-76	-26	20	87
unambiguous /s/	-84	-58	-30	38	76

This negative result carries weight only if listeners in fact restored the fricative consonants intended by the original speaker. Subjects' identification responses suggested that this was generally the case, particularly in the lexically unambiguous stimuli. Unambiguous /ʃ/ words were identified 97% correct when a vowel followed the initial consonant, but only 76% correct when /r/ followed; the errors in the latter case were predominantly F. Unambiguous /s/ words were identified 95% correct, regardless of following context. Ambiguous /ʃ/ words were correctly identified on 82% of the trials (errors being nearly always S), and ambiguous /s/ words on 84% (errors being predominantly SH, and sometimes TH). These percentages indicate that phoneme restoration did occur, and that it was guided both by residual acoustic cues and by a lexical bias.

GENERAL DISCUSSION

The present experiments addressed two conflicting hypotheses concerning the processes underlying phoneme restoration. The hypotheses rest on the basic assumption that an abstract phonological representation composed of phonemes can be distinguished from a concrete auditory percept composed of phones. Given this assumption, the *top-down completion hypothesis* stated that phoneme restoration occurs at the phonological level as a consequence of bottom-up and top-down constraints, but is not accompanied by active phone restoration from the acoustic signal. The phone is restored "in the mind's ear," if at all. Consequently, an extraneous noise replacing a fricative noise should be perceived veridically. The *segregation hypothesis*, on the other hand, stated that the acoustic material occurring in the place of the fricative noise is partitioned into a speech and a nonspeech portion, either by means of schema-guided auditory scene analysis (auditory induction) or by usurpation of the acoustic signal by a specialized speech module (pre-emptiveness). Consequently, the extraneous noise replacing [s] frication should be perceived as lower in pitch than an identical noise replacing [ʃ] frication or a stop closure. A probe noise matching task was employed to test these predictions.

Experiments 1 and 2 compared the perception of target noises replacing either [s] frication or a stop consonant closure silence. The results seemed to favor the segregation hypothesis: Noises replacing [s] were judged to be lower in pitch than noises replacing /t/ silence, relative to an external probe noise. However, the effect was much smaller than it would have been if full restoration of the [s] noise had occurred. Experiments 3 and 4, despite using rather similar stimulus materials, did not replicate the effect. One difference was that the /t/ release burst was removed from the control stimuli, in which the extraneous noise replaced the closure silence of /t/. Another difference was that the subjects were not as certain about the initial consonant as in the previous experiments, due to inclusion of nonword stimuli such as "teminar" and "selephone." Both these changes relaxed the tight constraints on phoneme identity and brought the phone restoration paradigm somewhat closer to the standard phoneme restoration situation, where no residual cues are present in the signal and a large variety of stimuli is employed. Experiment 5 indeed used a large sample of words and compared the perception of target noises replacing /s/ or /ʃ/. As in Experiments 3 and 4, no segregation effect was obtained.

The apparent segregation effect in Experiments 1 and 2 could have been due to an undesired effect of acoustic context in the control stimuli: The /t/ release burst, a brief noise strong in high spectral frequencies, may have been integrated with the target noise into a single noise percept, which naturally sounded higher in pitch. Alternatively, the conjunction of residual signal cues and lexical certainty may indeed have enabled listeners to carry out a segregative process, though it was not very effective. If the process occurs only under such highly constrained conditions, which are not representative of the typical phoneme restoration experiment, then its significance would seem to be limited. Yet, it might provide an existence proof of schema-guided influences on primitive auditory scene analysis (Bregman, 1990), or even of pre-emptiveness of a speech module (Mattingly & Liberman, 1990). This author's inclination is to consider the segregation effect in Experiments 1 and 2 a stimulus artifact.

In conditions that were still quite constrained in comparison to other phoneme restoration experiments and that did lead to restoration of the correct fricative consonant on most trials (Experiments 3-5), there was no trace of any segregative or pre-emptive processes. This negative result supports the top-down completion hypothesis. According to this hypothesis, the target noise aided phoneme restoration by masking the absence (or the presence, as the case may be) of a fricative noise, but it was not partitioned into the missing phone and an acoustic residue. Listeners appeared to hear the full target noise as an extraneous sound.

The great unknown in phoneme restoration experiments is what the restored speech really *sounds* like. The problem is that subjects' pronouncements about what they are "hearing" cannot be taken at face value; normally, they will be referring to a linguistic abstraction, not to a concrete auditory percept. The auditory percepts accompanying some phonemes (such as stop consonants) are difficult to describe and may be inaccessible to consciousness. For this reason, the present study employed fricatives, whose auditory correlates are relatively accessible. Repp (1981b) demonstrated that, in a fricative discrimination task, naive listeners seem to make judgments on the basis of their phonemic percepts; after some training, however, they are able to direct their attention to the auditory percepts of the fricative noises and judge their timbre quite accurately (presumably without losing the phonemic percept). Fricative consonants, therefore, can be perceived at two levels, one abstract and synthetic, the other concrete and analytic. It made sense, therefore, to ask whether the concrete percept is restored along with the abstract one in the phone restoration paradigm.

If the concrete fricative noise were restored in a top-down fashion, as a consequence of phoneme restoration, one might expect the illusory restored phone to have the prototypical auditory quality associated with the restored phoneme. Indeed, Warren and his collaborators have always claimed that a restored phoneme is auditorily indistinguishable from the original. Informal observations during the present studies suggested that the restored /s/ was subject to quality variations that depended on the acoustic information provided in its place. Obviously, these quality variations are related to the fact that the target phoneme was not always identified correctly. The impression of poor quality may have derived from either a lack of confidence in making the phonemic decision or from an awareness of the atypical auditory quality of the noise substituting for the fricative phone (or of its insufficiency as a potential masker of a real [s] noise, which amounts to the same thing). It is noteworthy that, in earlier phoneme restoration studies, the subjects were never asked to identify the restored phoneme or to comment directly on the quality of its sound or articulation. Instead, they were either required to localize the extraneous noise with respect to the speech (Warren's paradigm) or to discriminate "added" from "replaced" items (Samuel's paradigm). It is likely that subjects' above-chance discrimination performance in Samuel's experiments was due to their ability to distinguish quality differences in the critical phoneme.

One reason for the presence of these quality variations may lie in the *intensity* of the replacement sounds used.[6] In Warren's experiments, these sounds were always much louder than the speech sounds they replaced. Warren (1984) stressed that the replacement sound must be an *effective masker* of the speech sound for complete restoration to occur. If the extraneous sound masks the speech completely, then "replaced" and "added" conditions are, obviously, indistinguishable. Samuel's discrimination paradigm, however, required that "replaced" and "added" versions of a stimulus be discriminable to some extent, so that variations in the "strength" of phonemic restoration (as measured by d') as a function of other factors could be investigated. Therefore, Samuel used extraneous sounds that were similar to the critical speech segment in amplitude, and Samuel and Ressler (1986) introduced the use of signal-correlated noise, which retains the exact speech amplitude envelope. These replacement sounds did not mask the speech completely: In speech, spectral energy is unevenly distributed, whereas in signal-correlated noise the same total energy is spread over the whole frequency range. The peaks of the speech spectrum are therefore not masked by the noise.

The present study followed Samuel's methods, for a similar reason: If a very loud noise had been used to replace [s], as in Warren's experiments, any effect of perceptual segregation on its perceived timbre surely would have been too small to show up in the noise matching task. By using replacement noises whose intensity was comparable to that of the [s] noise to be restored, chances for observing a segregation effect were improved—or so it was thought. This expectation would hold only if segregation were guided by top-down expectancies, and a prototypical [s] spectrum were subtracted from the replacement noise. It appears, however, that this is not what happened; rather, the spectral quality of the perceived phone (if any) seemed to resemble that of the replacement noise. (No formal judgments of the sound quality of the fricative phone were obtained.)

It is possible that the present findings are specific to the stimuli used here. Not only were the replacement noises of the same intensity as the original fricative noise (as in Samuel's studies), but also they offered absolutely no bottom-up clues to spectral segregation, even when the original noise was part of the mixture, because of the unitary nature of random noise waveforms. It appears that auditory scene analysis cannot be driven entirely by top-down expectations, but the possibility remains that segregation processes triggered by the input are enhanced by top-down processes. It is relevant in that connection to note that the detectability of masked speech is not increased by knowledge of the speech to be detected, conveyed either through reading (Frost, Repp, & Katz, 1988) or lipreading (Repp, Frost, & Zsiga, in press). Also, lipreading information does not enhance phoneme restoration in Samuel's paradigm; in fact, it seems to interfere (Trout & Poser, 1990). These negative findings suggest that Bregman (1990) is right in surmising that the processes of primitive auditory scene analysis are autonomous.

The findings of Whalen and Liberman (1987) and Bentin and Mann (1990) provide a challenge (see Bregman, 1990, 1991, for discussion). They showed that a frequency transition embedded in speech contributes to speech intelligibility at intensities far below those at which the transition becomes detectable as a whistle or chirp. It seems as if the speech system appropriated most of the transition's energy, leaving only a residue for the auditory system. However, the finding may also be understood as showing that speech is perceived holistically (cf. Warren, 1983), whereas analytic perception of a speech component as "standing out" requires bottom-up cues to source segregation, such as a raised intensity. The fact that a transition segregated in this manner continues to contribute to speech perception can be understood as an instance of self-masking; that is, any auditory event may contain a weaker copy of itself. This is very similar to the situation in fricative perception, where the unmodified noise may be heard as a separate source with primitive auditory qualities (Repp, 1981b), without any apparent effect on perception of the fricative phoneme.

Two possible criticisms of the present research need to be addressed. One is that the subjects were simply unable to detect the target noise in the speech and responded on the basis of the probe noise only. While this seems a possibility with regard to Experiments 3 (which used only following probes) and 5 (which used a single target noise), other experiments showed that listeners were sensitive to differences among target noises varying in spectrum. More importantly, subjects never complained that they were unable to perform the task because they heard no noise in the sentence. There is little doubt that they did perform a noise comparison task, even though the timbre of the target noise was not easy to judge. The other criticism is that the stimuli were not conducive to phoneme restoration: The target noise served as a poor fricative noise, and subjects perceived whatever phoneme was most compatible with the acoustic evidence. This may describe the situation quite accurately. However, there is a continuum of possible scenarios between two extremes: on the one hand, Warren's classic demonstration in which the extraneous noise is not speechlike and is more intense than the speech signal it replaces, and, on the other hand, the original, unmodified speech signal (in which the extraneous noise is identical with the phone it replaces). It is not clear where phoneme restoration ends and ordinary phoneme perception begins. Restoration essentially refers to the apparent intactness of the speech, and more narrowly to the perception of the phoneme intended by the speaker. It is difficult to draw a line that excludes phoneme perception from the original signal as a legitimate instance of "restoration." Therefore, any objection to the present experiments (and to Samuel's similar paradigm) on the grounds that they did not lead to phoneme restoration is void. Moreover, the present Experiments 3-5, in contrast to most earlier experiments, actually provided evidence that the restored phonemes corresponded to those intended by the speaker, most of the time anyway.

Applying the notion of restoration to the extreme situation of self-replacement (i.e., an unmodified speech signal) is instructive with regard to evaluating the hypothesis of *phone* restoration, which was under test here. It would be bizarre to suggest that the original [s] noise needs to be segregated into two identical copies to support both phoneme and timbre perception. Clearly, it is just a matter of attending to one or the other of its two aspects: the concrete, auditory percept or the abstract, phonological one (for which it constitutes only one of a number of information sources). If an original [s] noise is replaced by a noise with different spectral properties (and this includes any noise obtained by adding some noise to an original [s] noise, for two added noises are perfectly fused and form a single sound), listeners will presumably adopt the same two perspectives, but with different degrees of emphasis: The less similar the replacement noise is to an [s] noise (or any other plausible fricative noise) in spectrum and intensity, the less it will contribute to phoneme perception, and the more readily it will be perceived as an extraneous sound. The present experiments apparently represented situations in which the replacement noise was sufficiently similar to [s] to contribute to phoneme identification, but also sufficiently dissimilar to be perceived as extraneous without much attentional effort. In Warren's experiments, the intense replacement noise presumably did not contribute to phoneme identification, which was entirely top-down, and was perceived as strikingly separate from the speech.

Phonological perception does not require awareness of specific auditory components. Some phones can be attended to as sound objects if they are in the acoustic input; if they are masked, they become indefinite. Restored phonemes are no less real if they have no accessible auditory substrate; after all, stop consonants are perceived this way all the time. Fricatives are special because their principal auditory component is relatively accessible as an independent percept. It seems, however, that fricatives can be perceived just as well if their characteristic noise is inaccessible or truly absent, as in the sine-wave speech stimuli of Remez and Rubin (1990), as long as other bottom-up and top-down constraints are sufficient for phoneme perception. Phone restoration may not occur because perception of auditory segments as such is not a prerequisite for phoneme perception. This latter point, of course, has long been made by Liberman and his associates (see, e.g., Liberman & Mattingly, 1989) and, in different ways, by Warren (1983) and Bregman (1990) as well. Nevertheless, confusion is perpetuated by frequent references to phonemes as "speech sounds" in the literature, as if they needed to be heard. The belief that restored phonemes are "heard" is closely connected to this usage. Acceptance of phonemes as abstract linguistic units without specific auditory properties may go a long way towards clarifying the intriguing phenomenon of phoneme restoration.

REFERENCES

Bentin, S., & Mann, V. A. (1983). Selective effects of masking on speech and nonspeech in the duplex perception paradigm. *Haskins Laboratories Status Report on Speech Research, SR-76*, 65-85.

Bentin, S., & Mann, V. A. (1990). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America, 88*, 64-74.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Bregman, A. S. (1991). The compositional process in cognition with applications to speech perception. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 197-223). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32*, 19-31.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27*, 143-165.

Foss, D. J., & Blank, M. A. (1980). Identifying the speech codes. *Cognitive Psychology, 12*, 1-31.

Frost, R., Repp, B. H., & Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory and Language, 27*, 741-755.

Glave, R. D. (1973). *Untersuchungen zur Tonhohenwahrnehmung stochastischer Schallsignale*. Hamburg, FRG: Helmut Buske Verlag.

Hesse, H.-P. (1982). The judgment of musical intervals. In M. Clynes (Ed.), *Music, mind, and brain* (pp. 217-226). New York: Plenum Press.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243*, 489-494.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14*, 211-235.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics, 28*, 213-228.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America, 69*, 548-558.

Mann, V. A., & Soli, S. D. (1991). Perceptual order and the effect of vocalic context on fricative perception. *Perception & Psychophysics, 49*, 399-411.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 71-102). Cambridge, MA: MIT Press.

Mattingly, I. G., & Liberman, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function* (pp. 775-793). New York: Wiley.

Mattingly, I. G., & Liberman, A. M. (1990). Speech and other auditory modules. In G. M. Edelman, W. E. Gall, & Cowan, W. M. (Eds.), *Signal and sense: Local and global order in perceptual maps* (pp. 501-520). New York: Wiley.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. (1983). The role of "chirp" identification in duplex perception. *Perception & Psychophysics, 33*, 469-474.

Remez, R. E., & Rubin, P. E. (1990). On the perception of speech from time-varying acoustic information: Contributions of amplitude variation. *Perception & Psychophysics, 48*, 313-325.

Repp, B. H. (1981a). On levels of description in speech research. *Journal of the Acoustical Society of America, 69*, 1462-1464.

Repp, B. H. (1981b). Two strategies in fricative discrimination. *Perception & Psychophysics, 30*, 217-227.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (pp. 243-335). New York: Academic Press.

Repp, B. H. (1989). Phone restoration. In T. Szende (Ed.), *Proceedings of the Speech Research '89 International Conference* (pp. 238-241). Budapest: Linguistics Institute of the Hungarian Academy of Sciences.

Repp, B. H., Frost, R., & Zsiga, E. (in press). Lexical mediation between sight and sound in speechreading. *Quarterly Journal of Experimental Psychology*.

Samuel, A. G. (1981a). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*, 474-494.

Samuel, A. G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1124-1131.

Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language, 26,* 36-56.

Samuel, A. G. (in press). A further examination of attentional effects in the phonemic restoration illusion. *Quarterly Journal of Experimental Psychology.*

Samuel, A. G., & Ressler, W. H. (1986). Attention within auditory word perception: Insights from the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 70-79.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 43,* 1735-1736.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America, 70,* 976-984.

Trout, J. D., & Poser, W. J. (1990). Auditory and visual influences on phonemic restoration. *Language and Speech, 33,* 121-136.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167,* 392-393.

Warren, R. M. (1976). Auditory illusions and perceptual processes. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 389-417). New York: Academic Press.

Warren, R. M. (1983). Multiple meanings of "phoneme" (articulatory, acoustic, perceptual, graphemic) and their confusions. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (pp. 285-311). New York: Academic Press.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin, 96,* 371-383.

Warren, R. M., & Obusek, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics, 9,* 358-362.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science, 176,* 1149-1151.

Warren, R. M., & Sherman, G. L. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics, 16,* 150-156.

Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237,* 169-171.

FOOTNOTES

* *Perception & Psychophysics,* in press.

[1] Warren (personal communication) does not see the two hypotheses contrasted here as mutually exclusive. He seems to reject the notion of an abstract level of phonological representation, at least as a possible site of phonemic restoration; instead, he regards syllables and words as holistic *auditory* patterns (cf. Warren, 1983). The distinction between concrete auditory and abstract schematic representations—mistaken though it may be—is fundamental to the present argument, however.

[2] The pre-emptiveness hypothesis was developed in connection with a task that required spectral completion of an isolated speech signal, while phonemic restoration requires temporal completion and is contextually guided. Possibly, therefore, the pre-emptiveness hypothesis, as conceived by Liberman and his colleagues, does not apply here.

[3] Due to characteristics of the original [s] noise, the "happiness" noises were lower in overall intensity than the "seminar" noises and had a somewhat irregular amplitude envelope. These amplitude differences cannot be seen in Figure 1, which shows amplitude-normalized plots of the spectra. (It would have been desirable to include the original [s] spectra in the figure, but they could not be located.)

[4] In addition to these main findings, the ANOVA revealed a number of significant interactions that are of little theoretical interest and whose discussion is omitted to ease the burden on the reader.

[5] One subject's comment suggests a possible explanation: "Selephone" is rather similar to the real word "cellophane" and may have activated that lexical entry.

[6] I am grateful to Richard Warren (personal communication) for drawing my attention to this important point.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 171-180

Tongue-Twister Effects in the Silent Reading of Hearing and Deaf College Students*

Vicki L. Hanson,† Elizabeth W. Goodell,†† and Charles A. Perfetti†††

To investigate whether deaf readers use phonological information during sentence comprehension, deaf and hearing college students performed a semantic acceptability task on tongue-twister and control sentences. Indicative of phonological coding, subjects' responses were influenced by the phonetic content of the sentences they were reading and by the phonetic content of a concurrent memory load task. That is, the subjects in both groups made more errors in their acceptability judgments when reading tongue-twister than when reading control sentences. In addition, subjects in both groups made more errors when the tongue-twister sentences and concurrent memory load numbers were phonetically similar than when they were phonetically dissimilar. These results support theories that assign phonological processes an important role in reading.

The question addressed by the present study is whether deaf readers use a phonological code during reading. Our interest in this question is prompted by the fact that deaf readers provide a strong test case for the importance of phonological processes in reading. These readers lack direct (i.e., auditory) access to the spoken English and, as a result, should have considerable difficulty in acquiring a phonological code. Moreover, deaf readers might use alternatives to phonological coding (e.g., visual or sign coding) if these alternatives were effective. The deaf subjects in the present study were all prelingually, profoundly deaf college students from deaf families, and had learned American Sign Language (ASL) as a first language. A finding of phonological coding[1] by these subjects, because of the degree of their hearing loss and because of their access to coding alternatives, would provide support for theories that place importance on phonological processes in reading (see for example, Huey, 1908/1968; Perfetti, 1985; Shankweiler & Crain, 1986).

We know from short-term memory research that it is possible for prelingually, profoundly deaf individuals to acquire and use a phonological code. For example, in the recall of printed words some deaf college students are sensitive to rhyme, performing more poorly when recalling lists of rhyming than nonrhyming words (Engle, Cantor, & Turner, 1989; Hanson, 1982; Hanson & Lichtenstein, 1990). Moreover, we know that the use of phonological coding in such short-term memory tasks is characteristic of deaf good readers, but not deaf poor readers (Conrad, 1979; Hanson, Liberman, & Shankweiler, 1984; Hanson & Lichtenstein, 1990). Although this short-term memory research may be used to argue that phonological processes underlie skilled reading even in deaf populations, there has been no convincing research that phonological coding actually is used by these individuals during reading.

Previous research directed at this question generally has used proofreading tasks. In proofreading tasks, subjects read passages of text and cancel out all the instances of a particular target letter in the text. Corcoran (1966) found that hearing

---

This research was supported, in part, by Grant NS-18010 from the National Institute of Neurological and Communicative Disorders and Stroke to Vicki L. Hanson while at Haskins Laboratories. In supporting this work, we are grateful to Drs. Michael Karchmer and Patrick Cox of Gallaudet University. We also wish to thank Deborah Kuglitsch for her help with testing subjects and Michael Studdert-Kennedy for his thoughtful comments on earlier versions of this manuscript.

subjects fail to detect more silent letters (such as the letter e in the word *tape*) than pronounced letters (such as e in the word *red*) in proofreading, and interpreted this finding as evidence of phonological coding during reading. Chen (1976) found no significant difference in the number of silent and pronounced *es* detected by congenitally deaf subjects, and concluded that these subjects, therefore, were not using phonological coding when reading. Similarly, Locke (1978) found no difference in target letter detections by deaf children when target letters occurred in words as a phonemically modal (e.g., *g* in *rag*) or nonmodal (e.g., *rage*) pronunciation. From this outcome, he also concluded that his deaf subjects were not using phonological coding when reading. It is now known, however, that performance in proofreading tasks is influenced by factors such as target position within a word, word frequency, and letter positional frequency (Frith, 1979; Smith & Groat, 1979). The stimuli of Chen (1976) and Locke (1978) were not controlled for these factors. With position of a target in a word controlled, Dodd (1987) found a difference in target letter detections by deaf children as a function of pronunciation. With word frequency controlled, Quinn (1981) similarly found a difference in target letter detections by deaf children as a function of pronunciation. None of these proofreading studies with deaf subjects has controlled for letter positional frequency in words, however. This orthographic structure variable is known to influence the reading of deaf subjects (Hanson, 1986). Due to such stimulus confoundings, the results from these proofreading studies with deaf subjects are inconclusive.

To avoid such stimulus confoundings, we, instead, used a "tongue-twister" task to examine the question of phonological coding by deaf readers. Hearing readers find it more difficult to read tongue-twister sentences (such as *The tired dentist dozed, but he drilled dutifully*), than syntactically and semantically matched controls (such as *The ill fireman tripped, and he walked unsurely*). For example, silent reading of tongue-twister sentences has been found to take longer than silent reading of the matched controls (Haber & Haber, 1982; see also Ayres, 1984). Similarly, semantic acceptability judgments of tongue-twister sentences have been found to be slower and less accurate than judgments of matched controls (McCutchen, Bell, France, & Perfetti, in press; McCutchen & Perfetti, 1982)

Some investigators have hypothesized that the tongue-twister effect is due to phonetic interference (Haber & Haber, 1982; McCutchen et al., in press; McCutchen & Perfetti, 1982). Other investigators, however, have questioned this interpretation of the tongue-twister effect. It has been suggested that the longer response times to tongue-twister sentences may be due to grapheme repetitions, rather than due to phoneme repetitions (Baddeley & Lewis, 1981). This issue of grapheme repetitions proved problematic in an earlier attempt to use the tongue-twister effect to examine reading by deaf adults. In that study, Treiman and Hirsh-Pasek (1983) found longer response times by deaf subjects when reading tongue-twister sentences than when reading control sentences. Their tongue-twister sentences, however, used repeated graphemes. This confounding of visual and phonetic similarity forced those investigators to regard their results as inconclusive with respect to phonological coding.

A study by McCutchen and Perfetti (1982), however, partially unconfounded visual and phonetic similarity in tongue-twister sentences by examining the tongue-twister effect with mixed-grapheme stimuli. These investigators used both same grapheme tongue-twisters, such as *The talented teenager took the trophy in the tournament,* and *mixed-grapheme tongue-twisters*, such as *The taxis delivered the tourists directly to the tavern.* In this latter example, the alveolar stop tongue-twister contained word initial phonemes of both /d/ and /t/. McCutchen and Perfetti (1982) found the tongue-twister effect in their study to be as large with the mixed-grapheme tongue-twisters as with same-grapheme tongue-twisters.

These mixed-grapheme stimuli reduced, but did not eliminate, the visual and phonetic confounding inherent in tongue-twister sentences. To eliminate this confounding in a tongue-twister task, McCutchen et al. (in press) looked for specific interference in semantic acceptability judgments as a function of the phonetic content of a concurrent memory load. Hearing subjects judged the semantic acceptability of alveolar fricative tongue-twisters (e.g., *The sparrow snatched the spider swiftly off the ceiling*) and alveolar stop tongue-twisters (e.g., *The taxis delivered the tourists directly to the tavern*). Before the subjects saw the sentences, five numbers appeared on the screen for subjects to memorize. The numbers began with alveolar stops (e.g., 12, 2, 20, 25, 22) or alveolar fricatives (e.g., 17, 6, 65, 16, 77). The subjects rehearsed the numbers, judged whether or not the sentence that followed was semantically acceptable, and then recalled as many numbers as they could.

McCutchen et al. found that rehearsal of alveolar stop and fricative numbers interacted with tongue-twister type. Namely, when the memory load consisted of numbers starting with a fricative, it took longer to respond to fricative sentences. When the memory load consisted of numbers starting with a stop, it took longer to respond to stop sentences. Recall of the numbers showed the same interaction, with more errors when the numbers and the sentences shared initial phonemes. The memory load numbers were visually distinct from each other (e.g., 7, 6, 16, 77) as well as from the sentence graphemes, demonstrating that the interference in this task was not visual. Rather, this interaction between tongue-twister type and memory load type could be clearly attributed to specific interference between the phonetic content of the sentences and the memory load.

The present study uses the procedures of McCutchen et al. (in press) to ascertain whether deaf college students use a phonological code in the silent reading of sentences. A phonological code would be indicated by tongue-twister effects and, more convincingly, by specific interference due to the phonetic content of the concurrent memory load.

Method

Subjects. The deaf subjects were 16 paid volunteers from Gallaudet University. All had deaf parents and were profoundly deaf, with hearing losses of 85 dB or greater, better ear average, as indicated by college records. All reported that ASL was used in the home, and that they considered ASL to be their first language. The data for two of these subjects were eventually dropped due to their performance on the number recall task. One subject produced no correct responses on the number recall task. The other subject failed to write down numbers on some trials. As a result, this subject lost his place on the answer sheet, making his number recall data impossible to score.

Reading scores were available for 11 of the remaining subjects from the comprehension test of *Gates-MacGinitie Reading Tests - Second Edition* (1978, Level F, Form 2). The median reading level for these subjects was grade 8.7 (range = 12.9+ to 3.3). Given that the average reading level of profoundly deaf high school graduates is approximately third grade (Conrad, 1979; Karchmer, Milone, & Wolk, 1979), most of these subjects were quite good deaf readers. Speech intelligibility scores were provided by Gallaudet University for those 12 students who had been tested by the Audiology Department at the University. The scores are based on a scale from 1 to 5, in which "1" is easily intelligible and "5" is unintelligible. For these subjects, one had speech rated a "2," another one had speech rated a "3.5," five had speech rated a "4," and five had speech rated a "5." On the whole, therefore, the speech of these subjects was not very intelligible.

The hearing subjects were 16 undergraduate students at the University of Connecticut who participated in the experiment to fulfill a course requirement.

Stimulus materials. The stimuli were the 144 alveolar stop, alveolar fricative, and control sentences previously used by McCutchen et al. (in press). These stimuli consisted of three sets of 48 syntactically matched sentences, differing in the initial consonants of the content words. The tongue-twister stimuli included both same-grapheme tongue-twisters and mixed-grapheme tongue-twisters. The content words in the alveolar stop sentences began with either /t/ or /d/, while the content words in the fricative sentence with either /s/ or /z/. The content words in the control sentences contained a mix of initial phonemes that excluded the two stop and fricative phonemes used in the experimental sentences.

McCutchen et al. constructed the semantically unacceptable sentences so that content words were rearranged within each sentence type. This allowed syntactic structure to remain unchanged. Half of the stimuli were semantically acceptable sentences, while the other half were semantically unacceptable. To ensure that the tongue-twister sentences were not more bizarre than the control sentences, McCutchen et al. (in press) asked hearing undergraduates to rate the semantic acceptability of all sentences designated as semantically acceptable. Results indicated that the tongue-twister sentences were no more or less bizarre than the control sentences. See McCutchen et al. (in press) for further details on stimulus selection. Examples of semantically acceptable and unacceptable sentences for fricative, stop, and control sentences are as follows:

Fricative sentences
Acceptable: The spacious zoo sits beside a sandy seashore.
Unacceptable: The salty zone smashed beside the skillful station.
Stop sentences
Acceptable: The tiny toddler dreamed of her toy tiger.
Unacceptable: The damaged detective dreamed in the tattered toddler.

Control sentences
Acceptable: The amusement park was beside a rocky beach.
Unacceptable: A black bush knew beside the walking stories.

Two test lists were created. These lists differed in terms of which sentences were presented in the memory load condition: the sentences tested in the memory load condition in one list were tested in the no memory load condition in the other, with the order in which the sentences appeared differing in the two lists. Equal numbers of each sentence type were included in each test list. The memory load condition preceded the no memory load condition on both lists.

There were two sets of numbers for the number recall task. In one set, the names for the numbers began with the alveolar stop /t/ (e.g., 2, 12, 22, 23, 24, 25, 28, 29) while in the second the names began with the alveolar fricative /s/ (e.g., 6, 7, 16, 17, 66, 63, 65, 68, 74, 79). A subset of five numbers was randomly picked for use on each recall trial.

Procedure. The subjects were individually tested. For the deaf subjects, the experimenter was a deaf native signer of ASL who was a former Gallaudet University student. For the hearing subjects, the experimenter was a graduate student at the University of Connecticut.

To familiarize themselves with the recall task, subjects began a testing session with four practice trials on which only the number recall task was given. These numbers were random, beginning not necessarily with either stops or fricatives. In this phase, subjects were instructed that their task was simply to recall a set of numbers briefly shown on a computer screen. They were to write this set of numbers in the spaces provided on their answer sheets when the numbers disappeared from the computer screen. The numbers appeared for 5 s, then disappeared.

Following completion of the number recall practice, the experimental session began. This session consisted of 4 practice trials, followed by testing with one of the two stimulus lists. Half of the subjects in each group saw each test list. The subjects were given a brief rest after every 24 test trials.

For testing with a memory load, subjects were instructed that they were to read each sentence, decide whether it made sense, indicate their decision by pressing a telegraph key, and then write the numbers. They were asked to sit with their index fingers resting one on each of two labelled telegraph keys, and indicate their decisions about semantic acceptability by pressing one of these two keys as quickly and as accurately as possible. The right-hand key was used to indicate semantically acceptable sentences.

The start of a trial was signalled by the appearance of five numbers at the center of the screen. After a blank interval of 250 ms, the sentence was then presented in the center of the screen, wrapped to a second line when too long to fit on one line. The sentence remained in view either until the subject pressed a response key or until 5 s had elapsed. The subjects were informed that feedback would be given on each trial. The feedback was the subject's RT (in ms) for his/her acceptability judgment. If the subject had failed to respond within the 5 s time limit, the words TOO SLOW appeared as feedback. (To anticipate, however, there were no cases in which a subject failed to respond during the time limit on test trials). The feedback, displayed for 250 ms, was centered six lines below the sentence.

If subjects asked how to know if a sentence "made sense," they were told to decide based on whether the meaning of the sentence was bizarre. They were informed that the spelling and syntax of sentences would always be correct, so that they need not look for spelling or grammatical errors.

At the end of the 72 experimental trials with the memory load, the answer sheets for number recall were collected and the subjects then performed 72 trials without the memory load. The procedure for the trials without a memory load was identical, except that no number string was presented before the test sentences and no number recall was required. No practice trials were given for this condition. The instructions for these trials without a memory load were given following the trials with a memory load. These instructions informed the subjects that they were simply to continue to decide whether the sentence made sense and that they no longer needed to recall numbers. As in the memory load condition, the subjects were given a brief rest after every 24 trials.

Results

The data for semantic acceptability judgments and number recall were analyzed to determine if performance on tongue-twister sentences differed from that on control sentences, and to determine whether there was any specific interference in sentence reading or number recall as a function of the phonetic content of the concurrent memory load. For the semantic acceptability judgments, median response times (RTs) for correct trials and mean percentage errors were analyzed. For number recall, subjects' mean percentage of errors in recalling the memory load numbers was analyzed.

As in the study of McCutchen et al. (in press), the number recall data were scored irrespective of order. Unless otherwise specified, only effects that reached statistical significance in both the subjects and the items analyses are reported.

Tongue-Twister Effects

ANOVAs were performed on the semantic acceptability judgment data for the factors of sentence type (tongue-twister sentences, control sentences), memory load (no memory load, memory load), acceptability (acceptable, unacceptable), and group (hearing, deaf). These analyses revealed significant tongue-twister effects in the error data [$F(1,28) = 14.40$, $MS_e = 73.60$, $p < .001$, for subjects; $F(1,140) = 4.78$, $MS_e = 456.40$, $p < .05$, for items]. Table 1 gives the mean errors and RTs for hearing and deaf subjects in this task, showing that subjects made more errors on tongue-twister (20.6% errors) sentences than on control sentences (16.3% errors).

The analyses of the semantic acceptability data further revealed that the deaf subjects were faster [$F(1,28) = 4.65$,$MS_e = 1937163.20$, $p < .05$, for subjects; [$F(1,139) = 368.05$, $MS_e = 118497.59$, $p < .001$, for items], although less accurate [$F(1,28) = 12.02$, $MS_e = 524.09$, $p < .005$, for subjects; [$F(1,140) = 14.24$, $MS_e = 262.40$, $p < .001$, for items], than the hearing subjects. The mean RT for the group of deaf subjects was 2578 ms, with 23.9% errors overall, as compared with a mean RT of 2966 for the group of hearing subjects, with 13.6% errors overall. In the RT analysis there was a main effect of acceptability, [$F(1,28) = 5.78$, $MS_e = 49343.45$, $p < .05$, for subjects; $F(1,139) = 4.73$, $MS_e = 160773.72$, $p < .05$, for items], reflecting faster RTs on semantically acceptable than unacceptable sentences.

For the number recall error data summarized in Table 2, ANOVAs were performed on the factors of sentence type (tongue-twister, control), acceptability (acceptable, unacceptable), and group (hearing, deaf). Only the main effect of group, $F(1,28) = 30.82$, $MS_e = 457.77$, $p < .001$, for subjects; [$F(1,140) = 339.33$, $MS_e = 86.98$, $p < .001$, for items], and the interaction of acceptability × group, [$F(1,28) = 14.73$, $MS_e = 21.36$, $p < .001$, for subjects; $F(1,140) = 7.90$, $MS_e = 86.98$, $p < .01$, for items], reached statistical significance in both subjects and items analyses. The main effect of group reflected more errors by the deaf subjects than by the hearing subjects.

Table 1. *Mean correct response times (RTs) and mean percentage errors for acceptability judgments in the No Memory Load and Memory Load conditions on tongue-twister and control sentences.*

	No Memory Load			
	RTs		Errors	
	Hearing	Deaf	Hearing	Deaf
Acceptable sentences				
Tongue-twister	2897 (654)	2587 (547)	14.0 (11.3)	22.8 (9.9)
Control	2698 (505)	2454 (561)	5.1 (7.3)	16.5 (9.2)
Unacceptable sentences				
Tongue-twister	2899 (542)	2568 (549)	12.7 (13.5)	26.5 (16.7)
Control	2859 (661)	2543 (531)	9.3 (12.5)	28.0 (22.1)
	Memory Load			
Acceptable sentences				
Tongue-twister	3158 (535)	2632 (619)	20.6 (13.5)	24.0 (6.0)
Control	2999 (410)	2471 (639)	15.2 (13.9)	14.6 (10.0)
Unacceptable sentences				
Tongue-twister	3169 (411)	2708 (547)	17.1 (12.6)	29.0 (17.0)
Control	3047 (611)	2656 (753)	14.8 (20.9)	29.6 (16.9)

Note. Standard deviations are shown in parentheses.

The interaction indicated that while the deaf subjects made somewhat more errors on tongue-twister than control sentences whether semantically acceptable or unacceptable, the hearing subjects made considerably more errors on the tongue-twisters when the sentences were unacceptable than when they were acceptable.

Table 2. *Mean percentage errors in recall numbers of the concurrent memory load on Tongue-Twister and Control sentences.*

	Hearing	Deaf
Acceptable sentences		
Tongue-twister	31.0 (12.1)	56.2 (10.2)
Control	30.3 (11.1)	55.1 (11.6)
Unacceptable sentences		
Tongue-twister	36.3 (10.3)	54.1 (9.6)
Control	32.5 (14.5)	51.6 (10.9)

Note. The standard deviations are shown in parentheses.

Specific Interference

To test for specific interference due to memory load, the acceptability judgment data and the number recall data of the Memory Load Condition were analyzed in terms of the factors of tongue-twister type (fricative, stop), memory load type (fricative, stop), acceptability, and group.

In the analyses of the acceptability judgments, specific interference effects were obtained in the error data. As shown in Table 3, the interaction of tongue-twister type × memory load type [$F(1,28) = 37.91$, $MS_e = 115.46$, $p < .001$, for subjects; $F(1,92) = 8.79$, $MS_e = 365.14$, $p < .005$, for items] reflected specific interference of the memory load. With fricative numbers as the memory load, there were more errors on acceptability judgments for fricative sentences (27.0%) than for stop sentences (18.9%), $t(29) = 3.50$, $p < .005$, two-tailed. Conversely, with stop numbers as the memory load there were more errors on acceptability judgments for stop sentences (25.6%) than for fricative sentences (17.0%), $t(29) = 3.38$, $p < .005$, two-tailed. Important for indicating phonological coding for deaf subjects, the specific interference interaction was present in the data of the deaf subjects alone [$F(1,13) = 39.17$, $MS_e = 100.18$, $p < .001$, for subjects; $F(1,92) = 7.90$, $MS_e = 384.35$, $p < .01$, for items]. The comparable analysis of the hearing subjects' data reached significance only in the subjects analysis, [$F(1,15) = 6.94$, $MS_e = 128.69$, $p < .02$, for subjects; $F(1,92) = 2.66$, $MS_e = 235.11$, $p < .11$, for items].

Table 3. *Mean correct response times (RTs) and mean percentage errors for acceptability judgments in the Memory Load condition as a function of Tongue-Twister Type (Fricative, Stop) and Memory Load Type (Fricative, Stop).*

	Fricative Memory Load			
	RTs		Errors	
	Hearing	Deaf	Hearing	Deaf
Acceptable sentences				
Fricative	3205 (651)	2512 (672)	26.1 (18.1)	29.7 (16.1)
Stop	3222 (525)	2537 (590)	19.7 (23.6)	16.7 (12.9)
Unacceptable sentences				
Fricative	3056 (540)	2484 (502)	16.8 (19.2)	36.9 (23.8)
Stop	2997 (491)	2726 (547)	17.7 (15.3)	21.5 (20.0)
	Stop Memory Load			
Acceptable sentences				
Fricative	3076 (614)	2727 (817)	13.6 (12.4)	14.4 (14.4)
Stop	3129 (630)	2755 (631)	21.9 (24.9)	34.5 (17.8)
Unacceptable sentences				
Fricative	3238 (608)	2769 (697)	12.5 (17.7)	28.5 (15.0)
Stop	3388 (389)	2854 (816)	19.9 (16.3)	27.4 (17.9)

Note. Standard deviations are shown in parentheses.

As in the tongue-twister analyses, the specific interference analyses of the semantic acceptability judgments revealed that the deaf subjects were faster, [$F(1,28) = 6.98$, $MS_e = 2083185.68$, $p < .05$, for subjects; $F(1,88) = 62.39$, $MS_e = 245496.05$, $p < .001$, for items], although less accurate [$F(1,28) = 5.37$, $MS_e = 657.66$, $p < .05$, for subjects; $F(1,92) = 8.26$, $MS_e = 254.32$, $p < .01$, for items] than the hearing subjects.

In the ANOVAs on the percentage of errors in number recall, no specific interference effects emerged. The only effects involving memory load type to reach statistical significance were the main effect of this variable [$F(1,28) = 62.48$, $MS_e = 209.47$, $p < .001$, for subjects; $F(1,88) = 160.24$, $MS_e = 77.63$, $p < .001$, for items], and a two-way interaction of memory load type × group [$F(1,28) = 12.12$ $MS_e = 209.07$, $p < .002$, for subjects; $F(1,88) = 32.65$, $MS_e = 62.23$, $p < .001$, for items]. As shown in Table 4, the main effect of memory load type was reflective of fewer recall errors when the memory load was stop numbers than when it was fricative numbers. The interaction of memory load type × group indicated that while the deaf subjects had roughly comparable accuracy in recall for fricative and stop numbers the hearing subjects had much better recall for fricative than stop numbers.

As in the tongue-twister analyses of the number recall data, the deaf subjects were less accurate than the hearing subjects [$F(1,28) = 31.95$ $MS_e = 862.91$, $p < .001$, for subjects; $F(1,88) = 365.39$, $MS_e = 62.23$ $p < .001$, for items]. The analysis of number recall also indicated an interaction of group × acceptability, [$F(1,28) = 16.48$, $MS_e = 48.85$, $p < .001$, for subjects; $F(1,88) = 5.60$, $MS_e = 62.23$, $p < .05$, for items], with the deaf subjects making roughly the same percentage of errors on both acceptable and unacceptable sentences, and the hearing subjects making nearly twice as many errors on acceptable than unacceptable tongue-twisters.

Discussion

The present study used a semantic acceptability task to examine whether deaf college students use phonological coding during reading. Indicative of phonological coding, subjects' semantic acceptability judgments showed both tongue-twister and specific phonetic interference effects.

Both the hearing and the deaf college students in this study made more errors on acceptability judgments when reading tongue-twister sentences than when reading control sentences. This outcome replicates the tongue-twisters effect from earlier work with hearing subjects, showing that tongue-twister sentences are more difficult to read than sentences that do not have this phonetic loading (Haber & Haber, 1982; McCutchen & Perfetti, 1982; McCutchen et al., in press). As noted previously, however, tongue-twister effects are often viewed skeptically due to their inherent confounding of visual and phonetic similarity. Even in studies, such as the present one, that attempt to reduce this confounding by using mixed-grapheme tongue-twisters, this confounding is not eliminated.

More conclusive evidence of phonological involvement in the present study was the specific phonetic interference. The error data for acceptability judgments showed more errors when the tongue-twister sentences and the concurrent memory load numbers were phonetically similar.

Table 4. *Mean percentage errors in recalling the numbers of the concurrent memory load.*

	Fricative Memory Load		Stop Memory Load	
	Hearing	**Deaf**	**Hearing**	**Deaf**
Acceptable Sentences				
Fricative	41.4 (18.1)	65.0 (11.0)	17.7 (10.7)	45.9 (14.2)
Stop	41.6 (17.2)	59.7 (14.3)	23.5 (14.0)	54.2 (12.0)
Unacceptable Sentences				
Fricative	46.0 (13.1)	58.8 (12.5)	23.3 (13.9)	50.9 (12.4)
Stop	48.3 (11.6)	53.8 (14.7)	27.7 (20.1)	53.0 (17.2)

Note. Shown are the percentage errors as a function of Tongue-Twister Type (Fricative, Stop) and Memory Load Type (Fricative, Stop).

Both the hearing and the deaf subjects had difficulty making acceptability judgments about fricative sentences when the memory load consisted of fricative numbers. Conversely, both groups had difficulty in making acceptability judgments about stop sentences when the memory load consisted of stop numbers. This outcome confirms the earlier finding of specific phonetic interference reported by McCutchen et al. (in press) for hearing college students, and extends the finding to deaf college students. Since the written forms of the sentences and numbers do not overlap visually, this result cannot be attributed to graphemic similarity. Rather, the interference would seem attributable to the phonetic similarity of the sentences and numbers.

There were no significant interactions involving subject group related to the tongue-twister effect or the specific interference effect. We note, however, a speed-accuracy tradeoff in the semantic acceptability judgment data, with the deaf subjects being faster and less accurate than the hearing subjects in their responding.

There was no evidence of either tongue-twister or specific interference in the number recall data. As noted earlier by McCutchen et al. (in press), there appear to be processing tradeoffs in this task, such that the tongue-twister and specific interference effects may appear either in the acceptability judgment data or the the number recall data. In the present study, these effects were apparent only in the acceptability judgments data. The number recall data showed only that the hearing subjects recalled more numbers than the deaf subjects, and that the hearing subjects were more influenced by the semantic acceptability of the sentences than were the deaf subjects.

The present results argue that a phonological code is indeed used in sentence comprehension during reading—by both hearing and deaf subjects. What role would it have? Most likely a phonological code is useful in maintaining words in working memory to allow for the processing of individual words and to assemble the words into phrases and sentences (Baddeley & Hitch, 1974; Huey, 1908/1968; Perfetti, 1985; Shankweiler & Crain, 1986). Research with deaf subjects has suggested that this code may be particularly useful for retaining words in their correct sequential order (Hanson, 1990; Lake, 1980) and for retention and interpretation of free morphemes and articles (Taeschner, Devescovi, & Volterra, 1988; Volterra & Bates, 1989).

A phonological code may not be the only code that can be used to mediate sentence comprehension during reading, however. For example, Treiman and Hirsh-Pasek (1983), testing deaf subjects who were less proficient readers than those of the present study, found that their subjects had more difficulty reading "finger-fumbler" (Klima & Bellugi, 1979) than control sentences. These finger-fumbler sentences had words whose corresponding signs were formationally similar. Those subjects, therefore, may have been using a sign code to mediate comprehension. It is noteworthy that the better deaf readers of Treiman and Hirsh-Pasek's (1983) study did not show finger-fumbler effects.

Strikingly, the deaf subjects of the present study all had deaf parents and reported ASL to be their first language. Despite this access to an alternative, these subjects used phonological coding in the present reading comprehension task. Also of interest is the fact that the deaf subjects of the present study did not tend to have outstanding speech production skills, as judged by listeners. Yet, again, even these subjects were found to use phonological coding. Importantly, these deaf subjects, as a group, had reading scores that were exceptionally high for deaf readers. It is possible, indeed likely, that we would not find evidence of phonological coding if a group of deaf poor readers were tested in this tongue-twister task. Such a finding would be consistent with short-term memory research, which has generally found the use of phonological coding to be limited to those deaf subjects who are the better readers (Conrad, 1979; Hanson & Lichtenstein, 1990).

Although a phonological code is often characterized as acoustic or auditory, our results suggest that these characterizations may be incomplete and, in the case of profoundly deaf persons, inaccurate. Although access to auditory information may enhance the effectiveness of a phonological code (Hanson, in press), research with deaf subjects suggests that descriptions of a phonological code should not be limited to being auditory in nature. Our results are consistent, for example, with the possibility that a phonological code for deaf readers could be *visually* derived, based on the look of words on the lips of speakers (see, also, Dodd & Hermelin, 1977). Indeed, the possibility that hearing persons also have a visual component to their phonological code is consistent with studies of lipreading and the "McGurk effect." In studies of lipreading, hearing subjects generally lipread as well as or better than deaf subjects trained in lipreading (Conrad, 1977; Pelson & Prather, 1974). The "McGurk effect" shows that in identifying certain phonemes, hearing subjects may report the phoneme they see on the lips of the speaker rather than the phonemes that they actually hear when the two sources of input conflict (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). Both the lipreading results and the "McGurk effect" indicate that hearing subjects, despite never having received explicit training in lipreading, know a great deal about the visual correlates of phonemes.

In summary, the tongue-twister and specific interference effects reported here provide evidence for the use of a phonological code in the silent reading of deaf college students. Given these subjects' access to coding alternatives and their difficulty in acquiring a phonological code, the present results provide strong support for theories that assign phonological processing an important role in reading.

REFERENCES

Ayres, T. J. (1984). Silent reading time for tongue-twister paragraphs. *American Journal of Psychology, 97,* 605-609.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation, Vol. 8.* New York: Academic Press.

Baddeley, A., & Lewis, V. (1981). Inner active processes in reading: The inner voice, the inner ear, and the inner eye. In A. M. Lesgold, & C.A. Perfetti (Eds.), *Interactive processes in reading.* Hillsdale, NJ: Erlbaum.

Chen, K. (1976). Acoustic image in visual detection for deaf and hearing college students. Journal of General Psychology, 94, 243-246.

Conrad, R. (1977). Lipreading by deaf and hearing children. *British Journal of Educational Psychology, 47,* 60-65.

Conrad, R. (1979). *The deaf schoolchild.* London: Harper & Row.

Corcoran, D. W. J. (1966). An acoustic factor in letter cancellation. *Nature, 210,* 658.

Dodd, B. (1987). Lip-reading, phonological coding and deafness. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading.* London: Erlbaum.

Dodd, B., & Hermelin, B. (1977). Phonological coding by the prelinguistically deaf. *Perception & Psychophysics, 21,* 413-417.

Engle, R. W., Cantor, J., & Turner, M. (1989). Modality effects: Do they fall on deaf ears? *Quarterly Journal of Experimental Psychology, 41A,* 273-292.

Frith, U. (1979). Reading by eye and writing by ear. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language, Vol. 1.* New York: Plenum Press.

Gates-MacGinitie Reading Tests (2nd ed., 1978). Boston: Houghton Mifflin.

Haber, L. R., & Haber, R. N. (1982). Does silent reading involve articulation? Evidence from tongue-twisters. *American Journal of Psychology, 95,* 409-419.

Hanson, V. L. (1982). Short-term recall by deaf signers of American Sign Language: Implications for order recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 572-583.

Hanson, V. L. (1986). Access to spoken language and the acquisition of orthographic structure: Evidence from deaf readers. *Quarterly Journal of Experimental Psychology, 38A,* 193-212.

Hanson, V. L. (1990, in press). Recall of order information by deaf signers: Phonetic coding in temporal order recall. *Memory & Cognition.*

Hanson, V. L. (in press). Phonological processing without sound. In S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman.* Hillsdale, NJ: Erlbaum.

Hanson, V. L., Liberman, I. Y, & Shankweiler, D. (1984). Linguistic coding by deaf children in relation to beginning reading success. *Journal of Experimental Child Psychology, 37,* 378-393.

Hanson, V. L. & Lichtenstein, E. L. (1990). Short-term memory coding by deaf signers: The notion of primary language coding reconsidered. *Cognitive Psychology, 22,* 211-224.

Huey, E. B. (1908/1968). *The psychology and pedagogy of reading.* Cambridge, MA: MIT Press.

Karchmer, M. A., Milone, M. N., Jr., & Wolk, S. (1979). Educational significance of hearing loss at three levels of severity. *American Annals of the Deaf, 124,* 97-109.

Klima, E., & Bellugi, U. (1979). *The signs of language.* Cambridge, MA: Harvard University Press.

Lake, D. (1980). Syntax and sequential memory in hearing impaired children. In H. N. Reynolds & C. M. Williams (Eds.), *Proceedings of the Gallaudet conference on reading in relation to deafness.* Washington, DC: Division of Research, Gallaudet College.

Liddell, S. K., & Johnson, R. E. (1986). American Sign Language compound formation processes, lexicalization and phonological remnants. *Natural Language and Linguistic Theory, 4,* 445-513.

Locke, J. (1978). Phonemic effects in the silent reading of hearing and deaf children. *Cognition, 6,* 175-187.

MacDonald, J. W., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24,* 253-257.

McCutchen, D., & Perfetti, C. A. (1982). The visual tongue-twister effect: Phonological activation in silent reading. *Journal of Verbal Learning and Verbal Behavior, 21,* 672-687.

McCutchen, D., Bell, L. C., France, I. M., & Perfetti, C. A. (in press). Phoneme-specific interference in reading: The visual tongue-twister effect revisited. *Reading Research Quarterly.*

McGurk, H., & MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Padden, C. & Perlmutter, D. (1987). American Sign Language and the architecture of phonological theory. *Natural Language and Linguistic Theory, 5,* 335-375.

Pelson, P. O., & Prather, W. F. (1974). The effects of visual message related cues, age and hearing impairment on speech reading performance. *Journal of Speech and Hearing Research, 17,* 518-525.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford.

Quinn, L. (1981). Reading skills of hearing and congenitally deaf children. *Journal of Experimental Child Psychology, 32,* 139-161.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition, 24,* 139-168.

Smith, P. T., & Groat, A. (1979). Spelling patterns, letter cancellation and the processing of text. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language, Vol. 1.* New York: Plenum Press.

Taeschner, T., Devescovi, A., & Volterra, V. (1988). Affixes and function words in the written language of deaf children. *Applied Psycholinguistics, 9,* 385-401.

Treiman, R., & Hirsh-Pasek, K. (1983). Silent reading: Insights from second-generation deaf readers. *Cognitive Psychology, 15,* 39-65.

BEST COPY AVAILABLE

Volterra, V., & Bates, E. (1989). Selective impairment of Italian grammatical morphology in the congenitally deaf: A case study. *Cognitive Neuropsychology*, 6, 273-308.

FOOTNOTES

[*] *Journal of Memory and Language*, *30*, (1991).

[†] IBM Research Division, Thomas J. Watson Research Center,Yorktown Heights, New York.

[††] Also University of Connecticut, Storrs. Now at the Infant Development Project, University of Washington, Seattle.

[†††] University of Pittsburgh, Pittsburgh.

[1] The term phonology has been used to describe signed languages as well as spoken languages. For example, this term has been used to describe the visible gestures of a signer's hands, face, and body that are the linguistic primitives of ASL (e.g., Liddell & Johnson, 1986; Padden & Perlmutter, 1987). In the present article, however, we use the term phonology to refer only to spoken languages.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 181-194

Whole Language vs. Code Emphasis: Underlying Assumptions and Their Implications for Reading Instruction*

I. Y. Liberman and A. M. Liberman

Promoters of Whole Language hew to the belief that learning to read and write can be as natural and effortless as learning to perceive and produce speech. From this it follows that there is no special key to reading and writing, no explicit principle to be taught that, once learned, makes the written language transparent to a child who can speak. Lacking such a principle, Whole Language falls back on a method that encourages children to get from print just enough information to provide a basis for guessing at the gist. A very different method, called Code Emphasis, presupposes that learning the spoken language is, indeed, perfectly natural and seemingly effortless, but only because speech is managed, as reading and writing are not, by a biological specialization, that automatically spells or parses all the words the child commands. Hence, a child normally learns to use words without ever becoming explicitly aware that each one is formed by the consonants and vowels that an alphabet represents. Yet it is exactly this awareness that must be taught if the child is to grasp the alphabetic principle and so understand how the artifacts of an alphabet transcribe the natural units of language. There is evidence that preliterate children do not, in fact, have much of this awareness; that the amount they do have predicts their reading achievement; that the awareness can be taught; and that the relative difficulty of learning it that some children have may be a reflection of weakness in the phonological component of their natural capacity for language.

Various studies have estimated the number of children who fail at reading to be 20-25 percent of the school population (Stedman & Kaestle, 1987). While it is generally agreed that this presents a serious problem, opinion is deeply divided about its underlying causes and inevitably, therefore, about the proper route to its solution. In this paper, we will explore two current views. One of these is commonly referred to by its partisans as Whole Language; the other, which we embrace, we call Code Emphasis, borrowing the name given it by Jeanne Chall (1967).

At the level of their most fundamental assumptions, Whole Language and Code Emphasis stand in stark contrast. Whole Language proceeds from the premise that learning to speak and learning to read are entirely comparable instances of language development. From this it follows that learning to read can and should be as effortless as learning to speak. Code Emphasis, on the other hand, recognizes that speech and reading must follow very different developmental paths. Speech is wholly natural, an integral part of the child's specialization for language. Because this specialization provides an automatic, tacit command of the complex relation between the sounds of speech and the abstract phonological structures they communicate, the acquisition of speech, whether in production or perception, is relatively effortless. A writing system, on the other hand, is an artifact, a biologically secondary code that maps to its natural language base in ways that must be quite consciously understood if it is to be used properly.

Accordingly, learning to read is, in the Code Emphasis view, a cognitive, intellectual achieve-

---

The preparation of this paper was aided by grants to Haskins Laboratories (NIH-NICHD-HD-01994) and to Yale University/Haskins Laboratories (NIH-21888-01A1). We are grateful to Anne Fowler and Donald Shankweiler for valuable criticisms. A special debt is owed to Alice Dadourian for her skill and infinite patience in helping us with the manuscript.

ment in a way that learning to speak is not. It is simply wrong to suppose, as Whole Language does, that they can be learned in the same epigenetic way.

What's Right With Whole Language

Some of what Whole Language espouses is undeniably right and inherently appealing, which may account for the wide currency it now enjoys. But this is so, in our view, only to the extent that it offers suggestions about instruction that sensible people like your grandmothers and ours would have regarded as truisms. Other, more fundamental, aspects of the Whole Language position are demonstrably false in ways that we will develop later.

But first let us consider some of the truisms. In a parent-teacher guide to the Whole Language movement that we take as our primary source (Goodman, 1986), Kenneth Goodman (past president of the International Reading Association as well as a founding father and leading writer in the Whole Language movement) makes numerous suggestions that no good teacher we know would quarrel with. For example, he says that we need committed teachers who will serve eagerly as "guides, facilitators, kid watchers" (p. 44), that in the preschool years children would profit from "literate environments with functional print everywhere" (that is, homes and schools with freely available books and magazines), as they would from "environments in which teachers and parents themselves enjoy reading, read to their children, take them to libraries, and generally expand the children's awareness of the functions of print" (p. 44).

The parent-teacher guide offers some equally unobjectionable ideas for the beginning literacy program. For example, it says that the best books for first graders are those with predictable stories that use frequent word repetitions, together with cyclical sequencing that provides lots of productive, self-motivating practice (p. 47). We cannot take exception to these notions either, except to remark that they are hardly original with Whole Language or with us, having been preempted by writers of children's stories from Mother Goose to Dr. Seuss.

The guide also recommends that reading and writing be integrated so that children can understand their reciprocal relationship (p. 47). It urges that schools should build on the language development children have attained before they reach school and expand it (p. 10). And it urges further that schools "respect learners: who they are, where they come from, how they talk, what experiences they had before coming to school" (p. 10).

There is a good deal more of this kind of advice in the parent-teacher guide. Though obvious, much of it nevertheless deserves more emphasis than it usually gets, and the Whole Language people are right to provide it. Surely, it is good to identify and promote practices that have been part of every good teacher's repertoire since the teaching of reading began.

What's Wrong With Whole Language

To find the important differences between Whole Language and Code Emphasis, one must put aside the easy truisms and look more deeply into the assumptions the two views make about the nature of language itself, and about the similarities or differences between the processes that underlie its spoken and written forms. What we see there, and what we mean now to say, is that the basic assumptions of Whole Language are wrong, and that they lead to recommendations about reading instruction that we consider grievously misguided. To avoid a common misunderstanding, we should recognize at this point that lots of children—perhaps as many as 75%—will discover the alphabetic principle, which is what they must understand if they are to read, no matter how unhelpful the instruction. But we find it ironic that, in order to succeed, these children might have to prevail over the misunderstandings of their teachers. Would even they not be better served by instruction that is designed to teach them what they need to know? In any case, we are left with the 20-25% who will *not* discover the point of the alphabet except as it is made apparent to them by appropriate instruction.

So let us examine a basic assumption of Whole Language, as exemplified by Goodman's description of a paradox he sees in the contrast between the ease of learning to speak and the relative difficulty of learning to read. He reminds us that infants learn to speak a language in a very short time and without formal training, but, when they reach school age, many of them begin to have difficulty with the kind of language development that Goodman associates with learning to read (p. 7). In his view, infants are "good at learning language when they need it to express themselves, as long as they are surrounded by people who are using language meaningfully and purposefully." As he sees it, children would be just as good at learning to read if only the task were made similarly meaningful and purposeful. Unfortunately, according to Goodman, teachers make learning to read difficult, "by breaking whole (natural) language into bite-size, abstract little pieces." School traditions, he says, "took apart the language into words, syllables, and isolated sounds." And this "postponed its natural purpose—the communication of meaning—and turned it into a set of abstractions unrelated to the needs and experiences of the child we sought to help" (p. 7).

That may sound plausible, but in our view it could hardly be more wrong. We agree, of course, that most children quickly master speech well enough to communicate usefully, and that, typically, they do this without explicit instruction, while, in contrast, many fail to learn to read and write. But before we accept Goodman's explanation that the schools are at fault by making a naturally easy task difficult (p. 8), we should ask two questions: first, is reading, like speaking, really all that natural; and, second, if not, what more is required if the child is to read and write?

To understand that the answer to the first question is an unqualified 'no,' one has only to consider such obvious facts about language as the following:

1 All communities of humans have a fully developed spoken language, but only a minority of these exists in a written form. Where there is a written form, many competent speakers do not, and, indeed, cannot use it effectively, no matter how strong the pressure to do so.

2 In the history of the race, as in the development of the child, speech comes first, reading second. Apparently, speech is as old as the human species, having evolved with it as perhaps the most important of its species-typical characteristics; alphabets, on the other hand, are developments of the last three or four thousand years, and they are cultural achievements, not the primary products of biological evolution.

3 Reflecting biological roots that run deep, speech employs a single, universal strategy for constructing utterances. All languages form all words (hence all utterances) by combining and permuting a few dozen consonants and vowels, meaningless segments that we will sometimes refer to, loosely, as phonemes. On the other hand, scripts, being artifacts, choose variably from a menu of strategies. Some, like the one we use, represent the phonemes. Others represent the more numerous syllables or, as in the case of Japanese, moras. And still others, like the Chinese, take the considerably more numerous morphemes as their irreducible units.

4 In order to develop speech, the normal child need only be in an environment where language is spoken; reading, on the other hand, almost always requires explicit tuition.

Given these telling facts about the differences between speech and script, it has to be both wrong and misleading to suppose, as Whole Language seems to, that they are psychologically and biologically equivalent vehicles for language. And, surely, it must be equally wrong, though perhaps even more misleading, to conclude, then, as the parent-teacher guide does (p. 24), that learning to read is, or can be, as natural a part of language development as learning to speak. Surely, it is plain that speech is biologically primary in a way that reading and writing are not. Accordingly, we suppose that learning to speak is, by the very nature of the underlying process, much like learning to walk or to perceive visual depth and distance, while learning to read and write is more like learning to do arithmetic or to play checkers. Just because learning to speak and learning to read can both be viewed as forms of language development—in the vacuous sense that both reflect the effect of experience on language behavior—it simply does not follow, as Whole Language would have it, that they are therefore equivalent forms of development, or that they can be instructed by experience in the same natural, unconscious way.

And so we are brought to the second question: Given children who have learned to speak, typically without conscious awareness of the underlying linguistic structures and processes they naturally deploy, what more must they learn if they would exploit their already impressive command of the language for the purpose of reading and writing it by means of an alphabet?

To see most clearly what more the would-be reader must learn about language, we need, first, to appreciate the critical differences between language and other natural means of communicating meanings, because that is where the most important aspects of the disagreement between Whole Language and us have their roots, and where we find the basis for the very different ways they and we understand the reading problem. We begin, then, by reminding ourselves that consonants and vowels are the essential structural elements of the phonological component of the child's natural capacity for language. The obvious and critically important function of the phonology is to form meaningful words by combining and permuting its small inventory of meaningless and abstract units in the elegantly principled ways that linguists have been concerned to characterize. For our purposes, however, it is important to know about the phonological strategy only that it is crucial for language because it makes possible vocabularies tens or even hundreds of thousands of times larger than could ever be managed if, as in all nonhuman systems, the signal for each 'word' were holistically different from every other one (See Liberman & Studdert-Kennedy, 1978 for further discussion). Thus, the phonology is not merely an inventory of sounds, but a marvelous system that comprehends all the words of the language, including even those that have yet to be uttered. It relates sound to meaning in a way that makes it one of the two properties (syntax is the other) that permits languages to be open and generative—that is, to convey an indefinitely large number of meaningful messages, including novel ones.

Communication among nonhuman animals is different in a critically important way, for, so far as anyone has been able to determine, the natural animal systems have no phonology (nor do they have syntax, for that matter), and, as a consequence, their message-carrying potential is severely limited. Lacking the phonological structures that make lexical generativity possible, nonhuman animals can convey in their natural communication only as many word-meanings as there are distinctively different signals they can make and perceive, and that is, at most, a few dozen. Moreover, short of calling a convention and getting all the animals to agree on a hitherto unused signal-to-meaning link, there is no way the animals can communicate a novel message. Thus, in contrast to language, which is lexically open because word meanings are conveyed by arranging and rearranging meaningless signal elements, the nonhuman systems attach meaning directly to each element and are, as a consequence, tightly and irremediably closed.

We see, then, that language would pay a terrible price if it were not phonologically based. Perhaps it would be of some comfort to the Whole Language people that in such a nonphonological world there would be no 'bite-size, abstract little pieces' for teachers to break a word into, and so 'postpone its natural purpose—the communication of meaning.' Each word would be conveyed by an unanalyzable signal, so meaning would be conveyed directly, just as Whole Language seems to think it should be. Unfortunately, there would not be many words.

In telling its very different story about the human and nonhuman capacities for communication, Whole Language scants the phonological faculty, emphasizing, instead, that our communication system is as it is because, by comparison with other species, we have (1) more to say, (2) a greater capacity for using symbols, (3) a need for social interaction, and (4) intelligence. The clear implication is that it is primarily these factors that make it possible for the child, but not the monkey, to learn language either in its spoken or its written form.

But, surely, animals do have something to say, and the most superficial study of the science of ethology reveals that many of them say it unerringly, often, and insistently. Whether what they say is "symbolic" depends on just exactly what one means by a symbol. Is a bird behaving symbolically when it produces a so-called 'mobbing' call? Does this call "symbolize" the presence of a predator, as well as the bird's desire to enlist the cooperation of its fellows in protecting their common interests? As for social interaction, many animal species are dependent on it for their very survival, and the biologically necessary interaction is always maintained, as it is in us, by communication. This leaves intelligence, about which we would only offer the notion that, whatever intelligence is taken to be, there is no strong reason to suppose that language is its inevitable product. Indeed, the matter might better be put the other way around: we may be intelligent, or, at least, appear so, because, being endowed with the species-typical and biologically distinct devices of phonology (and syntax)—the devices that critically distinguish language from all other forms of natural communication—we can manage cognitive processes in ways that creatures not so endowed must find impossible. Seen this way, language is not so much a result of intelligence as it is a critically important tool that an intelligence can use. Of course, a person with low intelligence will find that much less to do with the tool, and his use of it will surely reveal the poverty of his cognitive resources. But intelligence does not *cause* language, and it is not a truly sufficient condition for learning to read. Surely, there are intelligent people in societies which boast no readers.

So, in its most fundamental assumptions about language and its expression in speech or script, we find the parent-teacher guide doubly mistaken; first, in its assertion that learning to read can be like learning to speak, and, second, in its assumptions about the conditions that cause either kinds of learning to occur. We take it as undeniable that learning to read is *not* like learning to speak. As for the conditions that underlie learning to speak, they are but two: being neurologically normal in the several aspects of the language faculty, and having a normal exposure to the mother tongue. Learning to read imposes further requirements, of which more later.

Being neurologically normal for language means, *inter alia*, that the child has a natural capacity—indeed, a positive affinity—for phonological structures. This is why children are, in George Miller's apt phrase, 'spontaneous apprentices' in the business of language (Miller, 1977), acquiring new words at such a phenomenal rate that by the time they are six years of age they command a vocabulary of 13,000 words, of which 7800 are root forms. One year later their vocabulary comprises 21,600 words (12,400 roots), and just one year after that it has grown to 28,300 words and 17,600 roots (Templin, 1957). Such prodigies of language development would not be possible if there were no phonological system, and if children did not have, by virtue of their biology, a tacit command of its underlying structures and mechanisms (see Studdert-Kennedy, 1987 & M. Y. Liberman, 1983 for pertinent discussions).

We take it as given, therefore, that in teaching children to read and write, our aim must be to transfer the wonders of phonology from speech to script. In our view, this can be done only if the child comes to understand the alphabetic principle, the insight that words are distinguished from each other by the phonological structure that the alphabet represents. Surely, this is the principle that links the less natural mode of written communication to its natural, spoken base, and so makes available to the reader-writer the ready-made phonological system that gives to speech the incalculable advantages it enjoys.

But why might it be hard for children to grasp the alphabetic principle and so gain access to a phonology they have already pretty well mastered? To answer that question, which takes us to the very heart of the reading problem, we must first bother to understand how the phonemes are produced and perceived in speech, for only then can we see precisely in how far these processes must be different in writing and reading.

Consider, then, that if, as in every language, all utterances must be formed by variously stringing together two or three dozen consonants and vowels, the strings must, inevitably, run to considerable lengths. As a practical matter, then, there must be a way of producing these strings at some reasonable rate. Speech (and language, for that matter) would be impossible if, instead of saying 'bag,' we could only say 'buh, a, guh'; for to say 'buh a guh' is not to speak but to spell. In that case, language production would be impossibly tedious. To get a feel for what speech perception would be like, have someone spell a Victorian novel to you, letter by painful letter. Not only would communication be slow, but it would likely overreach the limits of working memory, and so make sentence comprehension all but impossible.

Of course, if speech were a matter of saying 'buh a guh' in place of 'bag,' then, as in the nonphonological case described above, the fundamental assumption of Whole Language would be more nearly right: reading and writing would be no more difficult and no less natural than speaking and listening, for any child who could say a word would, *ipso facto*, know how to spell and read it. But there would be no language worth writing and reading. Speech and language became possible only because there evolved in speech a specialization for the rapid and effortless production and perception of phonological structures. We and some of our colleagues believe that the strategy underlying this specialization was to define the phonemes, not as sounds, but as motor control structures we choose to call gestures. Thus, the phoneme we write as 'b' is a closing and opening at the lips; the phoneme we write as 'm' is that same closing and opening at the lips, combined with an opening of the velum; and so forth. In fact, the gestures are far more complex and abstract than this (see, for example, Browman & Goldstein, 1985; Liberman & Mattingly, 1985; Liberman & Mattingly, 1989), but, for our purposes, the important consideration is only that the gestural strategy permits coarticulation. That is, it permits the speaker to overlap gestures that are realized by different organs of articulation (as in the case of lips and tongue in [ba]) and to merge gestures that are produced by different parts of the same organ (as in the case of the tip and blade of the tongue in [da]). The consequence is that people can, and do, regularly speak at rates of 10 to 20 phonemes per second, which is nearly an order of magnitude faster than they could otherwise manage. That phonological elements are best defined as gestures is an hypothesis. Coarticulation, on the other hand is a fact, an essential characteristic of every language in the world.

But coarticulation was not there for the taking. It required the evolution of special articulatory gestures that lend themselves to being coarticulated, and that have come to serve a linguistic purpose and no other. (Thus, they are not engaged in eating, swallowing, breathing, or worrying a blackberry seed with the tongue.) It also required the evolution of a special method of controlling and coordinating these gestures, for speech mechanisms must produce enough overlap and merging to make high rates possible, while yet preserving the small set of invariant phonological structures that form all words (Liberman & Mattingly, 1989).

The relevance of all this to our concerns becomes apparent when one considers that, like other biologically specialized processes, those that exploit coarticulation to produce phonological structures go on automatically, below the level of awareness. The obvious consequence is that, to speak a word, a person need not know how to spell it, or even that it can be spelled. Nor does he need to know what articulatory gestures to make or how to make them. The speaker need only think of the word; the speech specialization does all the hard work for him, automatically selecting and coordinating the linguistically significant gestures that form the appropriate phonological structure.

Coarticulation has important consequences for speech perception as well (and hence for the would-be reader), for it folds into a single segment of sound information about several successive phonemes, and so relaxes the constraint on rate of perception imposed by the temporal resolving power of the ear (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). This produces a very complex relation between the sound and the phonological structure it conveys, but this considerable complication causes the listener no trouble; he has only to listen, for his phonological specialization parses the signal automatically, recovering the several coarticulated gestures that produced it. Once recovered in this automatic, precognitive way, the phonological structure that uniquely specifies the word 'bag' is available for whatever further use is to be made of it. We see, then, that production and perception of speech is easy, not because its processes are simple—for they are, in fact, marvelously intricate—but because the underlying specialization is so wonderfully adapted to its complex task.

But why make such a fuss about how we produce and perceive phonological structures if our concern is with reading and writing and the difficulties that attend them? The point is obvious, or so it seems to us. Given the biological specialization we have described, there is ordinarily nothing in children's experience with speech that will acquaint them with the alphabetic principle—that is, nothing to make them aware that all words are specified by an internal phonological structure, the shortest elements of which are the phonemes that the letters of the alphabet represent. Thus, the speech specialization causes a word like 'bag' to be coarticulated into a single, seamless piece of sound, even though it comprises three discrete phonemes. Given the automaticity of that specialization, the constituent phonemes do not ordinarily rise to the level of awareness. Therefore, the beginning reader does not understand why a word like 'bag' should be represented by three letters, or why its spelling should differ from that of 'sag' in the first letter, from 'big' in the second, and from 'bat' in the third. Moreover, the problem may resist the most obvious solution, because there is simply no way the teacher can divide the sound of 'bag' so as to recover its three phonemes. Saying 'buh a guh' to the child does not necessarily help all that much, since 'buh a guh' is the wrong word. At all events, we can now see that the normal processes of speech not only fail to reveal the internal structures of words, but may, indeed, obscure them. Of course, the requisite awareness can be developed, as it obviously has been in all literate people. Indeed, developing that awareness should be the first aim of the reading teacher. But that takes some doing.

All that we have said about the beginning reader applies to the condition of our ancestors at the time they developed the alphabet. After all, human beings had been producing and perceiving speech for tens or hundreds of thousands of years before that moment, just three or four thousand years ago, when it occurred to someone that words did not differ holistically, one from another, but only in the particulars of their internal structure. Given that momentous, if seemingly simple, linguistic discovery, it only remained for someone to get the idea that if each phonological element were represented by an identifiable, but wholly arbitrary, optical shape, then all could read and write, provided only that they knew the language and were consciously aware of the internal structure of its words.

Once again the wrong assumptions by Whole Language about the nature of spoken and written language lead them to advice about instruction that is, in our view, likely to be unhelpful. Thus, they assert in the guide (p. 9) that one trouble with reading instruction as it is sometimes practiced is that it is designed to make the child into a linguist. This, they say, is entirely unnecessary—after all, the child need not be a linguist in order to speak—and even harmful, since it makes learning to read an intellectual, and therefore disagreeable, task. For exactly the reasons we've just given, Code Emphasis agrees that a child need not be a linguist in order to speak. But it holds that to use an alphabetic writing system properly, the child must be led to the same linguistic insight—and it was a linguistic insight—that underlay the development of the alphabet. Becoming enough of a linguist to appreciate that all words have an internal structure need not be a disagreeable task, as we will argue later, but, agreeable or not, it is a necessary achievement for anyone who would take advantage of the alphabetic mode for the purpose of reading and writing.

So much, then, for the differences between Whole Language and Code Emphasis in the assumptions they make about spoken and written language. We turn now, though only briefly, to an equally important difference between the two views in what they have to say about the nature of the reading process itself. Put with admirable succinctness by Goodman himself, the Whole Language assumption is that reading is a "psycholinguistic guessing game" (Goodman, 1976). By this, Goodman means that (presumably skilled) readers merely sample the print, apprehending some words and skipping others. Then, using their normal and natural language processes, they guess at the message by taking advantage of context, their knowledge of the world, or, indeed, anything else that will spare them the inconvenience of actually reading what the writer had, in fact, written. As we will see later, this leads the Whole Language people to advocate actually *teaching* the child to guess. But, for now, the point is simply that their fundamental assumption about skilled reading is contrary to fact. The elegant studies of eye movements during reading by Rayner and his associates have shown conclusively that good readers fixate and thus 'read' every word (Rayner & Pollatsek, 1987). It is only the poor readers who sample the print, picking out words here and there, and then guessing at the rest. Other studies by Perfetti's group in Pittsburgh show that truly literate people are much less likely to use a guessing strategy (Perfetti & Lesgold, 1979). And, finally, there is the demonstration by Gough and associates (Gough, Alford & Holley-Wilcox, 1981) that, more often than not, guessing from context leads to errors of a most egregious sort. Thus, their well-educated skilled reader, given appropriate context and unlimited time, correctly guessed only one word in four.

Relevant Research

We have pointed to differences between Whole Language and Code Emphasis in their general hypotheses about the nature and causes of reading disability, and we have said how well or how poorly those hypotheses accord with some of the most basic facts about the nature of spoken and written language. It is time, now, to make these hypotheses more specific and explicit, and to inquire, at least briefly, into the research by which they are to be evaluated. We would, of course, like to be even handed in this matter, offering data on both sides of the argument. Unfortunately for that purpose, Whole Language seems to have led to little, if any, relevant research. Based, as it is, on the assumption that children should learn to read just as they learned to speak, Whole Language cannot have thought it relevant to discover what is necessary for reading beyond the conditions that were sufficient for the development of speech. Yet obvious and undeniable facts about language tell us that something more *is* necessary, and common sense suggests that identifying this something provides the basis for knowing what a proper program of instruction must be designed to do. At all events, Whole Language has made no research contribution to this important issue. Nor have its partisans mounted studies aimed at finding out what distinguishes those children who become good readers from those who do not, or how to predict at an early stage which children are at risk of failure. In the case of the research that is relevant to one tenet of Whole Language—that is, the data we cited in connection with the claim that reading is psycholinguistic guessing—the outcome was, as we said, directly contrary to what Whole Language had, in the absence of any data, supposed to be the case. We therefore turn to the specific, testable questions about reading which Code Emphasis has led to, and to the research studies by which those questions have been, and are being, answered.

As we have said, our guiding premise is that proper use of an alphabetic script requires, most importantly, an awareness of the fact that words are specified by their internal phonemic structure and, further, that such awareness does not come for free. This has seemed to us quite obvious, given what we know about language and the way an alphabet conveys it. But it seems not to appear so to everyone—certainly not to the proponents of Whole Language—so we seek further support for it, not in anything more we might learn about language and alphabets—for the evidence there seems full and convincing—but rather more directly in the results of research on learning to read and on the difference between those who succeed and those who do not.

(1) *Is knowing a language sufficient to provide beginning readers with the phonological awareness*

(awareness of the internal structure of words) that they need if they are to apply the alphabetic principle? Phonological awareness means simply the more or less explicit understanding that words are made of discrete units—that a word like 'bag' has three such units, that 'brag' has four, and that 'brags' has five. It does not entail knowing how to spell a word, only that it *can* be spelled. On the basis of all the considerations about speech and the development of the alphabet that we summarized in the earlier parts of this paper, we many years ago (Liberman, 1973) assumed, as we have already said, that beginning readers would usually lack phonological awareness. Then, seeking direct support for our assumption, we tested it on preschoolers (Liberman, Shankweiler, Fischer, & Carter, 1974). Using a task that was presented (and accepted) as a tapping game, we found that only 17% of the kindergartners 'passed' according to any reasonable criterion of passing. (Many more—about 48%—performed well with syllables, which was to have been expected, since syllables, unlike phonemes, are always marked acoustically by a discrete, vocalic center.) Thus, we found that relatively few kindergartners are aware that words can be taken apart into units like those the alphabet represents. Following that study, many others, using different kinds of measures, have arrived at the same conclusion (for reviews, see Blachman, 1988; Routh & Fox, 1984). Indeed, research on illiterate adults has shown that they, too, are lacking in phonological awareness (Byrne & Ledez, 1983; Liberman, Rubin, Duques, & Carlisle, 1985; Lukatela, Liberman, & Shankweiler, in preparation; Marcel, 1980; Morais, Cary, Alegria, & Bertelson, 1979; Pratt & Brady, 1988; Read & Ruyter, 1985).

So it is now quite firmly established that neither experience with speech nor cognitive maturation is sufficient to acquaint a person with the principle that underlies all alphabets. As for the relatively few kindergartners who have the awareness that the use of an alphabet presupposes, we might guess that they have acquired it as a result of the kinds of more or less analytic linguistic activities they may have engaged in at home (rhyming and various linguistic games, for example) or even by observing the print of stories read aloud to them. In any case, the teacher should never assume that the beginning reader has achieved the cognitive insight about language that reading requires.

(2) *Are there individual differences in degree of phonological awareness that correlate with (are predictive of) reading achievement*? The answer to that question is presented very pointedly by Bryant and Goswami (1987), who say that "the discovery of a strong relationship between children's phonological awareness and their progress in learning to read is one of the great successes of modern psychology." The relevant evidence is most impressive, coming as it does from studies that covered a wide range of cultural and economic backgrounds, as well as a number of different languages: in English, (Blachman, 1984, Bradley & Bryant, 1983, Fox & Routh, 1980, Goldstein, 1976, Helfgott, 1976, Liberman, 1973, Mann & Liberman, 1984, Olson, Wise, Conners, & Rack, 1989, Treiman & Baron, 1981); in Swedish, (Lundberg, Olofsson, & Wall, 1980, Lundberg, 1988, Magnusson & Naucler, 1987); in French, (Bertelson, 1987; Morais, Cluytens, & Alegria, 1984); in Spanish, (de Manrique & Gramigna, 1984); in Italian, (Cossu, Shankweiler, Liberman, Tola, & Katz, 1988); in Serbo-Croatian, (K. Lukatela et al., in preparation). Indeed, several investigators have arrived at the conclusion that, of all possible tests, the kind that measures some aspect of phonological awareness is the best single predictor of reading achievement (Blachman, 1988; Golinkoff, 1978; Lundberg et al. 1980; Mann, 1984; Routh & Fox, 1984; Stanovich, 1985; Vellutino & Scanlon, 1987).

(3) *Can phonological awareness, as such, be taught to pre-readers, and, if so, does teaching it have consequences for later achievement in reading*? Appropriate methods of reading instruction are, by their very nature, likely to make learners sufficiently aware of phonological structure to allow them to appreciate, and hence to apply, the alphabetic principle. Whether or not the teacher is herself aware of the importance of phonological awareness, she is likely, given common sense and a reasonable approach to instruction, to call the children's attention to the internal structure of words and to how this structure is reflected in the alphabetic spelling. Indeed, it is presumably by this route that most of us learned what we needed to know. Moreover, as we pointed out earlier in this paper, some children will infer the principle, no matter how ill-advised the method of instruction. Still, it is important to know whether phonological awareness can be taught as a prelude to reading instruction, and whether such teaching has happy consequences, especially for children who are at risk. As for children at risk, Bradley and Bryant (1983) at Oxford University first identified them by appropriate measures of phonological ability at ages 4 and 5, and then showed that specific training in phonological

classification caused them to progress better in learning to read, as measured even four years later (Bradley, 1988), than control groups that had equivalent training in semantic (as opposed to phonological) classification, or that received no linguistic training at all. Subsequently, Bertelson's group in Belgium (Content, Kolinsky, Morais & Bertelson, 1986), Blachman in Syracuse (Ball & Blachman, 1988), Lundberg's laboratory in Sweden (Lundberg, Frost, & Peterson, 1988; Olofsson & Lundberg, 1983) and Vellutino and Scanlon (1987) in Albany have all achieved salutary effects with training of normal randomly selected groups of kindergartners.

There is, then, considerable research support for the conclusion that phonological awareness can be trained in the prereader and that such training causes children to make better and faster progress when they later undertake to learn how to read and to spell. This conclusion applies to normal children and also to children who, because of deficiencies in measurable aspects of phonological awareness, are presumably at risk.

(4) *Given the differences among children in the ease with which they can be made aware of phonological structure, differences that correlate so highly with achievement in reading, where do we look to find the source of these differences?* Together with our colleagues, we are currently pursuing the hypothesis that the source is the phonological component of the child's specialization for language. If that hypothesis is correct, then relative difficulty in achieving phonological awareness should be only one of several symptoms. Among other symptoms of a weak phonology, we should expect problems with verbal short-term memory, because, as is well known, (Baddeley, 1968; Conrad, 1964; Liberman, Mattingly, & Turvey, 1972), such memory requires the use of phonological structures. If those structures are weak, short-term memory should suffer We are not surprised, therefore, to find that poor readers do, in fact, perform more poorly than age-matched good readers on tasks requiring immediate memory of verbal items, though with nonverbal tasks, such as those that require memory for nonsense shapes or photographs of unfamiliar faces, the poor readers are not at a disadvantage (Gathercole & Baddeley, 1989; Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982; Mann & Liberman, 1984; Rapala & Brady, 1990; but see Pennington, 1989).

Also consistent with the assumption that poor readers suffer from a relatively weak phonology is the finding by Brady and associates (Brady, Shankweiler, & Mann, 1983) that poor readers need a higher quality of signal than good readers for error-free performance in the perception of speech, but not for the identification of nonspeech environmental sounds. The other side of this coin is the finding by Catts (1986) that reading-disabled junior high students made significantly more errors than matched normal pupils on demanding tasks of speech production (tongue twisters, polysyllabic words, and the like).

Even the well-documented naming problems of reading disabled second-graders—problems that, on their face, invite an interpretation that places the problem in the semantic component—turn out on investigation to lie in the phonology. Thus, in a series of carefully controlled studies with the Boston Naming Test, Katz (1986) found that poor readers who misnamed objects could nevertheless describe their functions accurately, recognize the correct names when given a series of choices, or even generate the correct name when given a phonological prompt. Apparently, their problem was that they could not access the right phonological structure, presumably because it was weakly established.

Though this issue is far from settled, we are encouraged to believe that individual differences in the ease with which children grasp the alphabetic principle originate in the phonological component of their language faculty, not in some cognitive (for example, analytic) capacity that cuts across linguistic and nonlinguistic domains (for a review of the evidence, see Wagner & Torgesen, 1987). Until we see compelling evidence to the contrary, we will, therefore, suppose that this is the right way to understand the child whose difficulties with reading do not extend to all intellectual tasks.

Instruction by Whole Language

Given the profound differences in the underlying assumptions of Whole Language and Code Emphasis, we should expect equally profound differences in the instructional methods they rationalize. In fact, these differences exist, though they are sometimes papered over by a throw-away line or two in the parent-teacher guide. Thus, the parent-teacher guide does say (p. 46) "Cultivate the alphabetic principle." But it does not tell parent or teacher just exactly how it is to be cultivated or why. Instead, it devotes considerable space to promoting procedures that seem designed, not to reveal the principle, but to obscure it. Thus, it says (p. 43) that "literacy development is a matter of learning to use just enough print, language structure, and meaning, and to keep it all in the proper personal and cultural perspective." We are not told exactly how the beginner arrives at the decision that he has used "just enough print" or how he uses "the proper personal and cultural perspective" to learn to read. In any case, the guide repeatedly leaves the teacher with the notion that the alphabetically represented word is somehow an encumbrance rather than a medium for conveying meaning. Thus, the guide says "Readers are seeking meaning, not sounds or words," (p. 38), but does not explain how one reaches the meaning without grasping the words.

The guide does permit a little leeway—it allows readers to "use their developing phonic generalizations to help when the going gets tough," but goes on to warn that "If they are lucky enough *not* to have been taught phonics in isolation, with each letter equally important, then they will *not* be diverted from developing the strategies necessary to select just enough graphic information to get to the sense they are seeking" (p. 38). The guide suggests that instead of bothering with all the graphic information, "learners need to know which available cues are most useful in a particular written context." Figure 1 displays an exercise seen on a poster board in a reading-methods classroom in a School of Education where the Whole Language approach was in operation. The exercise was presumably meant to develop just such guessing skills. We find shocking the message it appeared to convey—that learning to guess meaning from that most minimal of graphic cues, the initial consonant or two is a desirable route to reading acquisition.

In much the same vein, the guide explains that in its natural, easy-going classrooms, "trial and error risk-taking on the part of the learner is an absolute requirement" (p. 43). In the risk-taking, guessing approach, the learner necessarily makes errors. Not to worry. The parent-teacher guide assures us that learners' errors are accepted, indeed "celebrated," if they contribute to making sense: "No one is perfect," says the guide, "and sense, rather than error-free performance is the main point of reading" (p. 47). Therefore, we are not surprised to learn from graduate students taking a course in reading at a university where Whole Language was favored that they are told to reward children who, for example, read "Crest" instead of "toothpaste" in a story about brushing teeth. Crest is a toothpaste, after all, their professor reminded them. Similarly, we see a feature story in the *New York Times* (March 31, 1990) about an adult in a community literacy class who reads "kids" for "children." Presumably, Whole Language would find this quite acceptable, even praiseworthy. Children are kids, aren't they? Surely, this is wholly consistent with what we learn from the parent-teacher guide, where we are told that such errors may be viewed as "charming indications of growth toward control of language processes" (p. 19). We fear that they may serve, instead, as indicators that the children are still out of the loop (*pace* President Bush), wholly ignorant of what the game is about. They will probably not figure it out on their own, and their errors will become less and less charming as time goes on. We can only hope that when these readers graduate into the real world they will not skip the words they cannot read or substitute their own words in the instruction manuals they use to operate the machinery of the workplace.

Hey la la, Ho la la

My d ___ and I

Tr ___ to m ___

With ch ___ and p ___

Hey ___ ___, ___ ___ ___

Oh d ___ t ___ c ___

If y ___ ___ st ___

We'll n ___ g ___ th ___

Figure 1. Poster found in a reading clinic classroom in a School of Education, apparently used to train children in guessing from context.

Lest you suppose that the parent-teacher guide we have been quoting is not representative of the thinking of the Whole Language movement (even though it is written by their leading author), let us quote from a recent issue of the Whole Language Teachers Newsletter (Spring 1988) about what to do if the reader encounters an unfamiliar word (we are grateful to Professor Charles Read for calling our attention to this). The newsletter says:

> Foremost on the list of Don'ts are sound-it-out and look-for-familiar-word-parts-within-the-word because these activities divert the reader's attention from meaning...Good Things to Do include skip it, use prior information...read ahead, re-read, or put in another word that makes sense.

But all of this has got to be bad advice. For, surely, what the reader wants to get from the printed page is what the writer *actually* said, not what the reader thinks he *might* have said, given the reader's guess from context and his 'cultural and personal perspective.' Is it not just the most accomplished writers who, in an effort at precision and style, tend to avoid the inexact, soggy, and cliche-ridden prose that results when words are made to appear only in those contexts so very usual that guessing is, in fact, as good as reading? And is it not just that text which is designed to inform the readers, instead of repeating what they already know, that will necessarily lie outside their 'personal and cultural perspective.' If the readers are relying on what they already know, as Whole Language would have them do, how on earth are they ever going to use reading to learn something new? And how, if they follow the prescriptions of Whole Language, could they ever appreciate the beauty of poetry or the majesty of the Gettysburg Address? Are the readers to substitute their guesses for the words of Shakespeare and Lincoln? Are not the 'cultural and personal perspective' of the writer more important than those of the reader if the reader is to understand what the writer wrote?

We are here reminded of a pertinent example of how a dyslexic boy in the third grade, a child who had not mastered the alphabetic principle, attempted to follow a 'psycholinguistic-guessing-game' approach and the principles recommended in the Whole Language newsletter. (Though this episode occurred in the clinical practice of one of us some thirty years ago, she has never forgotten it.)

The child was given a passage to read, the first sentence of which was: "A boy said, 'Run little girl.' " The child knew how to read the word 'A' correctly, probably because it is the name of the letter. When he came to the word, 'boy,' he engaged in guessing and risk-taking as recommended in the guide and produced "baby," presumably on the basis of a cue available to him—the 'b.' Now he had 'A baby.' He didn't know 'said' so he looked ahead and found 'run,' which he did know. But what to do with the word between 'baby' and 'run'? Well, he used his "intact language skills" and figured that since there was an 's' in the word, and since he needed something to fill in there, he would say 'is running' because that would make sense syntactically (though it may be somewhat dubious semantically). But then he ran into a big snag. Unfortunately, 'little' was a word he knew by sight, so he was forced to read it correctly. Now his sentence said, "A baby is running little," and he was confronted with the word 'girl' that he did not know. At this point he gave up trying to make sense and simply threw in 'go' which he knew begins with the right letter. And his sentence now read "A baby is running little go." So much for the guide's assumption that "kids are universally able to sort out language as they use it to meet their functional needs." So much for children's ability to use their "developing phonics generalizations to help when the going gets tough" (p. 38) and so much for using "their prior learning and experience to make sense of the texts, guessing what will occur next." And so much for their ability "to select just enough graphic information to get to the sense [they are] seeking."

A three-step Whole Language procedure for introducing children to the reading process that appeared recently in an academic journal seems to rise above the print altogether, disregarding the information that letters, as opposed to pictures, are supposed to convey (Norris, 1989). In the first step, after reading the children the story of Goldilocks and the Three Bears, the teacher asks them to draw pictures of the story. We are told that the teacher then "reads the pictures" the child has drawn, saying "Once upon a time there were three bears—a papa bear, a mama bear, and a baby bear." In the second step, the teacher asks the children to add letters to the picture. The children, who apparently have not been taught much about letters, scrawl some marks on the paper. The teacher points to the marks while saying: "This says, 'Papa bear's porridge was too hot.' This word says, 'Mama bear's porridge is too hot.' " (In our day Mama bear's porridge was too cold and Baby bear's was just right—but this is apparently a different story.) Finally, the teacher asks the children to use the bears' names in their story and shows them how the names are represented in the book. The children copy some of the letters, apparently at random—in the figure accompanying the article the letters include a string of L's and T's and another of capital- and lower-case b's and some figures that look like 8's. The teacher points to the child's letters while saying: "You wrote about Baby Bear. Baby Bear's porridge was too hot." This sort of thing might be acceptable as a game for two and three-year-olds, but it is so misleading as to be questionable even for them. Those pictures the children drew do *not* say "Once upon a time there were three bears." The scrawls they made tell nothing about the heat of the porridge. Nor do the letters the children selected to copy from the printed story relate to the words in the story. Encouraging a child in this way to believe that they do may plant the false notion that print is like a picture.

Code Emphasis and the Alphabetic Principle

We have offered ample reason to believe that, by the very nature of the difference in underlying process, learning to read will not be so easy and effortless as learning to speak, whether the approach is Whole Language or Code Emphasis. But just because Code Emphasis makes intellectual demands of the child, it does not follow that its procedures must entail the mindless, sounding-out drills that the parent-teacher guide fears. In fact, Code Emphasis can be carried out in a pleasant, game-like atmosphere, with children participating happily, and with rapidly growing understanding of the alphabetic principle. Appropriate procedures have been described in detail elsewhere (e.g., Blachman, 1987; Elkonin, 1973; Engelmann, 1969; Gallistel, Fischer, & Blackburn, 1977; Liberman, 1989; Liberman, Shankweiler, Blachman, Camp, & Werfelman, 1980; Lindamood & Lindamood, 1975; Gillingham & Stillman, 1956; Rosner, 1975; Slingerland, 1971), and cannot, for want of space, be given here. Suffice it to say that not only can all these procedures be presented in ways that are enjoyable for the child, but they can also be effective, even in inner-city schools, as is strongly suggested by the dramatic increases in reading scores that were obtained after the procedures were introduced (Blachman, 1987; Enfield, 1987; Calfee, Lindamood & Lindamood, 1973; Wallach & Wallach, 1976; Williams, 1985).

The gains in reading in some of these studies were measured by determining how well a child could read isolated words (e.g., Calfee et al., 1973) but in others (Blachman, 1987; Enfield, 1987), standard tests of reading comprehension were used. Moreover, research evidence supports the intuitively obvious view that skill in comprehension is highly correlated with skill in decoding of single words (Curtis, 1980; Jastak & Jastak, 1978; Perfetti & Hogaboam, 1975; for a general discussion of this matter, see Gough & Tunmer, 1986). We remark on this, only because the Whole Language people so often suggest that Code Emphasis produces children who decode but don't comprehend. It is, of course, true as we have noted earlier, that children who are not adepts of decoding will invest so much time and attention in just getting the word that they will overreach the capacity of the working memory that is essential to processing the sentence (Perfetti & Lesgold, 1979). However, this is not a reason for encouraging them to skip or to fill the words in by guessing; rather it is a sign that they need, by practice, to make the decoding process automatic and relatively effortless. Even when children appear to have mastered the words, they may still have trouble with the sentence, not because they have bothered to read words, but because they have a deficiency in the syntactic component of their language faculty, independently of whether the constituent words are spoken or read. Indeed, there is evidence that poor readers do have relatively greater syntactic difficulties than good readers, even when the sentences are spoken to them (Smith, Macaruso, Shankweiler, & Crain, 1989), and it is interesting from our point of view that there are some reasons for supposing that such difficulties may, in fact, be traceable to the same phonological deficiencies that underlie the problems with decoding (Shankweiler, 1989). Finally, we offer the common-sense observation that comprehension will sometimes be impossible for reasons that have nothing to do with either reading or language ability. Thus, no matter how skillful or linguistically gifted, readers will, for example, not comprehend a paper on a scientific subject in a field utterly foreign to them. But whatever the reasons for failing to understand a text, none of them in any way suggests that the remedy is to teach the reader to guess at the words or to skip them. Given the nature of language, it is simply inconceivable that texts can be understood except by taking account of the words they comprise. Teaching the children what they need to know in order to read those words fluently must be the primary aim of reading instruction. What they need to know, and what their experience with language has not taught them, is no more and no less than the alphabetic principle.

REFERENCES

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology, 20,* 249-264.

Ball, E. W., & Blachman, B. A. (1988). Phoneme segmentation training: Effect on reading readiness. *Annals of Dyslexia, 38,* 208-225.

Bertelson, P. (1987). *The onset of literacy: Cognitive processes in reading acquisition.* Cambridge: MIT Press.

Blachman, B. A. (1988). Phonological awareness and word recognition: Assessment and intervention. In A. G. Kamhi & H. W. Catts (Eds.), *Reading disorders: A developmental language perspective.* San Diego, CA: College-Hill Press.

Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade achievement. *Journal of Educational Psychology, 76,* 610-622.

Blachman, B. (1987). An alternative classroom reading program for learning disabled and other low-achieving children. In W. Ellis (Ed.), *Intimacy with language: A forgotten basic in teacher education.* Baltimore: Orton Dyslexia Society.

Bradley, L. (1988). Rhyme recognition and reading and spelling in young children. In W. Ellis (Ed.), *Intimacy with language: A forgotten basic in teacher education* (2nd ed.). Baltimore: Orton Dyslexia Society.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature, 30* 419-421.

Brady, S. A., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35,* 345-367.

Browman, C. P., & Goldstein, L. M. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), *Phonetic linguistics.* New York: Academic Press.

Bryant, P. E., & Goswami, U. (1987). Beyond grapheme-phoneme correspondence. *Cahiers de Psychologie Cognitive, 7,* 439-443.

Byrne, B., & Ledez, J. (1983). Phonological awareness in reading disabled adults. *Australian Journal of Psychology, 35,* 345-367.

Calfee, R. C., Lindamood, P., & Lindamood, C. (1973). Acoustic-phonetic skills and reading: Kindergarten through twelfth grade. *Journal of Educational Psychology, 64,* 293-298.

Catts, H. W. (1986). Speech production/phonological deficits in reading disordered children. *Journal of Learning Disabilities, 19(8),* 504-508.

Chall, J. (1967). *Learning to read: The great debate.* New York: McGraw Hill.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75-84.

Content, A., Kolinsky, R., Morais, J., & Bertelson, P. (1986). Phonetic segmentation in prereaders: Effect of corrective information. *Journal of Experimental Child Psychology, 42,* 49-72.

Cossu, G., Shankweiler, D., Liberman, I. Y., Tola, G., & Katz, L. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psychology, 9,* 1-16.

Curtis, M. E. (1980). Development of components of reading skill. *Journal of Educational Psychology, 72,* 656-669.

de Manrique, A. M. B., & Gramigna, S. (1984). La segmentacion fonologica y silabica en ninos de prescolar y primer grado. *Lect. y Vida, 5,* 4-13.

Elkonin, D. B. (1973). U. S. S. R. In J. Downing (Ed.), *Comparative reading.* New York: Macmillan.

Enfield, M. L. (1987). A cost-effective classroom alternative to "pull-out programs." In W. Ellis (Ed.), *Intimacy with language: A forgotten basic in teacher education.* Baltimore: Orton Dyslexia Society.

Engelmann, S. (1969). *Preventing failure in the primary grades.* Chicago: Science Research Associates.

Fox, B., & Routh, D. K. (1980). Phonetic analysis and severe reading disability in children. *Journal of Psycholinguistic Research, 9,* 115-119.

Gallistel, E., Fischer, P., & Blackburn, M. (1977). *Manual: GFB sequence of objectives for teaching and testing reading in the concept transfer sequence.* Hamden, CT: Montage Press.

Gathercole, S. E., & Baddeley, A. D. (1989). The role of phonological memory in normal and disordered language development. In C. von Euler, I. Lundberg, & G. Lennerstrand, (Eds.), *Wenner-Gren Symposium Series 54, Brain and Reading.* London: MacMillan Press.

Gillingham, A., & Stillman, B. (1956). *Remedial training for children with specific disability in reading, spelling, and penmanship* (rev. ed.). Cambridge, MA: Educators Publishing Service.

Goldstein, D. M. (1976). Cognitive-linguistic functioning and learning to read in preschoolers. *Journal of Experimental Psychology, 68,* 680-688.

Golinkoff, R. M. (1978). Phonemic awareness skills and reading achievement. In F. B. Murray & J. H. Pikulski (Eds.), *The acquisition of reading: Cognitive, linguistic, and perceptual prerequisites* (pp. 23-41). Baltimore: University Park Press.

Goodman, K. S. (1976). Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading.* Newark, DE: International Reading Association.

Goodman, K. S. (1986). *What's whole in whole language: A parent-teacher guide.* Portsmouth, NH: Heinemann.

Gough, P. B., Alford, J. A., Jr., & Holley-Wilcox, P. (1981). Words and contexts. In O. J. L. Tzeng & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology.* Hillsdale, NJ: Erlbaum Associates.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6-10.

Helfgott, J. (1976). Phoneme segmentation and blending skills of kindergarten children: Implications for beginning reading acquisition. *Contemporary Educational Psychology, 1,* 157-169.

Jastak, J. F., & Jastak, S. (1978). *The wide range achievement test.* Wilmington, DE: Jastak Associates.

Katz, R. B. (1986). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. *Cognition, 22,* 225-257.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic reading in the beginning reader. *Journal of Experimental Child Psychology, 32,* 474-484.

Liberman, A. M. (1989). Reading is hard just because listening is easy. In C. von Euler, I. Lundberg, & G. Lennerstrand, (Eds.), *Wenner-Gren Symposium Series 54, Brain and Reading.* London: Macmillan Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 6, 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243,* 489-494.

Liberman, A. M., Mattingly, I. G., & Turvey, M. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes and human memory.* Washington: V. H. Winston and Sons.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology. Vol. VIII, Perception.* Berlin: Springer-Verlag.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society, 23,* 65-77.

Liberman, I. Y. (1989). Phonology and beginning reading revisited. In C. von Euler (Ed.), *Wenner-Gren International Symposium Series: Brain and reading* (pp. 207-220). Hampshire, England: Macmillan.

Liberman, I. Y., Mann, V., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex, 18,* 367-375.

Liberman, I. Y., Rubin, H., Duques, S., & Carlisle, J. (1985). Linguistic skills and spelling proficiency in kindergartners and adult poor spellers. In D. B. Gray & J. F. Kavanagh (Eds.), *Biobehavioral measures of dyslexia*. Parkton, MD: York Press.

Liberman, I. Y., Shankweiler, D., Blachman, B., Camp, L., & Werfelman, M. (1980). Steps toward literacy. In P. Levinson & C. Sloan (Eds.), *Auditory processing and language: Clinical and research perspectives*. New York: Grune and Stratton.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Liberman, M. Y. (1983). A linguistic approach. In M. Studdert-Kennedy (Ed.), *Psychobiology of language*. Cambridge, MA: MIT Press.

Lindamood, C. H., & Lindamood, P. C. (1975). *The A.D.D. program, Auditory discrimination in depth*. Hingham, MA: Teaching Resources.

Lukatela, K., Liberman, I. Y., & Shankweiler, D. (in preparation). *Cognitive characteristics of Yugoslav illiterate farm women*.

Lundberg, I. (1988). Lack of phonological awareness—a critical factor in developmental dyslexia. In C. von Euler, I. Lundberg, & G. Lennerstrand (Eds.), *Wenner-Gren Symposium Series 54, Brain and Reading*. London: Macmillan Press.

Lundberg, I., Frost, J., & Petersen, O.-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23/3*, 263-284.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years, predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

Magnusson, E. & Naucler, K. (1987). Language disordered and normally speaking children's development of spoken and written language: Preliminary results from a longitudinal study. *Report Uppsala University Linguistics Department, 16*, 35-63.

Mann, V., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities, 17*, 592-598.

Mann, V. (1984). Longitudinal prediction and prevention of reading difficulty. *Annals of Dyslexia, 34*, 117-137.

Marcel, A. (1980). Phonological awareness and phonological representation: Investigation of a specific spelling problem. In V. Frith (Ed.), *Cognitive processes in spelling*. London: Academic Press.

Miller, G. (1977). *Spontaneous apprentices*. New York: Seabury Press.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a consequence of phones arise spontaneously? *Cognition, 7*, 323-331.

Morais, J., Cluytens, M., & Alegria, J. (1984). Segmentation abilities of dyslexics and normal readers. *Perceptual Motor Skills, 58*, 221-222.

Norris, J. A. (1989). Facilitating developmental changes in spelling. *Academic Therapy, 251*, 97-108.

Olofsson, A., & Lundberg, I. (1983). Can phonemic awareness be trained in kindergarten? *Scandinavian Journal of Psychology, 24*, 35-44.

Olson, R., Wise, B., Conners, F., & Rack, J. (1989). Deficits in disabled readers' phonological and orthographic coding: Etiology and remediation. In C. von Euler, I. Lundberg, & G. Lennerstrand (Eds.), *Wenner-Gren Symposium Series 54, Brain and Reading*. London: Macmillan Press.

Pennington, B. F. (1989). Using genetics to understand dyslexia. *Annals of Dyslexia, 39*, 81-93.

Perfetti, C. A., & Hogaboam, T. (1975). The relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology, 67*, 461-469.

Perfetti, C. A., & Lesgold, A. M. (1979). Coding and comprehension in skilled reading and implications for reading instruction. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading*, Volume 1. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pratt, A. C., & Brady, S. (1988). Relation of phonological awareness to reading disability in children and adults. *Journal of Educational Psychology, 80(3)*, 319-323.

Rapala, M. M., & Brady, S. (1990). Reading ability and short-term memory: The role of phonological processing. *Reading and Writing: An Interdisciplinary Journal, 2*, 1-25.

Rayner, K., & Pollatsek, A. (1987). Eye movements in reading: A tutorial review. In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading*. London: Erlbaum.

Read, C., & Ruyter, L. (1985). Reading and spelling skills in adults of low literacy. *Remedial and Special Education, 6*, 43-52.

Rosner, J. (1975). *Helping children overcome learning disabilities*. New York: Walker and Co.

Routh, D. K., & Fox, B. (1984). "Mm... is a little bit of May: Phonemes, reading, and spelling." In K. D. Gadow & P. Bialen (Eds.), *Advances in learning and behavioral disabilities* (Vol. 3). Greenwich, CT: JAI Press.

Shankweiler, D. (1989). How problems of comprehension are related to difficulties in decoding. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle*. Ann Arbor: The University of Michigan Press.

Slingerland, B. H. (1971). *A multisensory approach to language arts for specific language disability children: A guide for primary teachers*. Cambridge, MA: Educators Publishing Service.

Smith, S. T., Macaruso, P., Shankweiler, D., & Crain, S. (1989). Syntactic comprehension in young poor readers. *Applied Psycholinguistics, 10*, 429-454.

Stanovich, K. E. (1985). Explaining the variance in reading ability in terms of psychological processes: What have we learned? *Annals of Dyslexia, 35*, 67-96.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language, perception and production* (pp. 67-84). London: Academic Press.

Stedman, L. C., & Kaestle, C. E. (1987). Literacy and reading performance in the United States from 1880 to the present. *Reading Research Quarterly, 22*, 8-46.

Templin, M. (1957). *Certain language skills in children*. Chicago: University of Chicago Press.

Treiman, R. A., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & R. G. Walker (Eds.), *Reading research: Advances in theory and practice, 3*. New York: Academic Press.

Vellutino, F. R., & Scanlon, D. (1987). Phonological coding and phonological awareness and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33/3*, 321-363.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Wallach, M. A., & Wallach, L. (1976). *Teaching all children to read*. Chicago: The University of Chicago Press.

Williams, J. P. (1985). The case for explicit decoding instruction. In J. Osborn, P. Wilson, & R. Anderson (Eds.), *Reading education: Foundations for a literate America*. Lexington, MA: D. C. Heath.

FOOTNOTE

*This paper is a revision of a talk given before the Orton Dyslexia Society in Dallas on November 29, 1989. It appears in *Annals of Dyslexia, 40*, (1990). For recent reviews that elaborate and document the views we here express, see: A. Liberman (1989) and I. Y. Liberman (1989).

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 195-208

Depth of Phonological Recoding in Short-Term Memory*

Yi Xu†

In testing the hypothesis that surface phonetic form is included in short-term memory (STM) representation, the tone sandhi phenomenon in Mandarin Chinese was exploited, and, as a prerequisite, the hypothesis that tonal similarity affects STM of verbal material in a tone language was also tested. In Experiment 1, subjects recalled visually presented sequences of seven monosyllabic Chinese morphemes having either the same tone or different tones. More errors were made on the monotonal sequences than on the multitonal sequences, confirming the effect of tonal similarity on STM. In Experiment 2, subjects recalled visually presented sequences of disyllabic nonsense words. The sequences were designed in such a way that half of them were subject to the tone sandhi rule in Mandarin Chinese whereas the other half were not. The consequence of applying the tone sandhi rule, as designed, was to make all the first characters in the sequence identical in pronunciation, thus creating potential phonological confusion. More errors, indeed, occurred on the sequences subject to the tone sandhi rule than on those not subject to it, indicating the existence of a surface phonetic representation in STM. The findings in this study provide further insight into the phonological mechanism of STM. Different accounts for this mechanism are also discussed in the light of those new findings.

INTRODUCTION

It has been long established that linguistic material, whether presented auditorially or visually, has to be retained in speech form in memory if it is to be remembered for more than a few milliseconds. This has been evident from the results of various studies showing the phonological nature of short-term memory (STM). It has been found that the errors that the subjects make in immediate recall are mostly phonological confusions; that is, items that share the same phonological features with the stimulus items were used in erroneous responses instead of items that share visual or semantic properties with the stimulus items (Brady, Mann, & Schmidt, 1987; Conrad, 1964; Wickelgren, 1965a, 1965b). It has also been found that subjects make more errors recalling series containing phonologically similar items than they do recalling phonologically dissimilar series. This is true for both auditorially presented sequences and visually presented sequences (Baddeley, 1966, 1968; Conrad, 1964). In visual presentation, the effect of phonological similarity has been found to be independent of the nature of the stimuli, as long as they can be represented linguistically. It has occurred with written letters (Conrad, 1964), numbers or words (Baddeley, 1966, 1968), and nameable pictures (Conrad, 1972).

Neither does the effect of phonological similarity seem to vary with the nature of the writing system. Erickson, Mattingly, and Turvey (1977) found that Japanese kanji characters were harder for native speakers to recall correctly when the set of characters to be remembered had phonetically similar readings. Tzeng, Hung, and Wang (1977) showed that phonological similarity, whether in the visually presented items to be recalled or in the shadowing items that were auditorially presented and repeated by the subjects, interfered with STM for Chinese characters. Yik (1978) showed that the effect of phonological similarity on STM of Chinese characters is larger than the effect of visual similarity, although the latter was also found to be significant.[1] Zhang and Simon

---

This research was supported by NICHD Grant HD-01994 to Haskins Laboratories. I thank Ignatius G. Mattingly, Alvin M. Liberman, Bruno H. Repp, Arthur S. Abramson, Diane Lillo-Martin and Robert G. Crowder for their valuable comments on an earlier version of this paper.

(1985) found that the STM span for Chinese radicals without common names was no greater than three items, and the STM span for complete homophones[2] was the same as that of radicals. But the STM span for non-homophones was significantly greater than that for homophones. Ren and Mattingly (1990) investigated STM for Chinese characters in relation to the reading ability of second-grade primary school children. They found that good readers were penalized more by phonological similarity in the series to be recalled than were poor readers. But visual similarity had no effect on either the good readers or the poor readers.

STM has also been shown to be sensitive to different aspects of phonology. Most of the studies showed rhyming effects upon STM (Baddeley, 1966; Erickson et al., 1977; Shankweiler, Liberman, Mark, Fowler, & Fisher, 1979). Consonantal similarity was also found to have a significant effect on STM (Brady et al., 1987), although some studies found that the effect of consonantal similarity is much smaller than that of vowel similarity (Tzeng et al., 1977).

The phonological nature of STM implies, as shown by some of the studies mentioned above, that if the material to be remembered is orthographic, it has to be first transformed into a speech-like form. This kind of transformation process is usually referred to as phonological recoding. The term phonological recoding is, however, ambiguous in that the levels of phonology involved are not specified. It could be understood as the underlying morphophonemic level, at which the phonological representation of a linguistic unit, (e.g., a morpheme or a word) is the one defined by the lexicon; it could be understood as the surface phonetic level, at which linguistic material is in a form similar to that generated in overt speech; or it could be understood as including both. To determine which level is involved in STM, it is necessary to find cases in which there are different representations for the linguistic material to be remembered at the two levels. The representations could be different because of differences in the specifications that are unique to either level. It has been shown by Baddeley, Thomson and Buchanan (1975) that when the number of syllables and the number of phonemes in words are held constant, STM span is inversely related to the temporal duration of the vowels in words, indicating a use of the surface specifications of speech as memory codes. The underlying and the surface forms could also be different because of discrepancies between the intended phonological representation and the actual surface sounds. Locke and Kutz (1975) showed that for children who mix certain contrasting sounds in their speech but correctly distinguish the same contrast in their perception, it is the surface sound they would produce rather than their phonological intention, that provides codes for their short-term memory.

Discrepancies between the morphophonemic representation and the surface phonetic representation of linguistic material are more commonly seen when smaller linguistic units are combined to form larger ones; for instance, when morphemes are combined to form words and when words are combined to form phrases and sentences. This combination often results in different number of distinctive categories at the two levels because of the application of phonological rules. This phenomenon can be employed to check whether surface phonetic representation or underlying morphophonemic representation is used as a code for STM. One way to do this is to find linguistic material that has more similarity in its surface forms than in its underlying forms and to employ the common finding that phonological similarity impairs the immediate recall of linguistic material. This is the approach used in the current research. This approach calls for the testing of linguistic material that contains more similarity in the surface phonetic representation than in the underlying morphophonemic representation. Specifically, the tone sandhi phenomenon in Mandarin Chinese was manipulated to create this material. Before presenting the details of the experiments, however, it is desirable to introduce some relevant background about the Chinese language family, as well as about Mandarin Chinese.

The Chinese languages are all tone languages. In a tone language, differences in the change of pitch provide not only intonational information about a sentence, e.g., whether it is a question or a statement, but also distinction between lexical items. In other words, two utterances otherwise identical in pronunciation might represent completely different lexical items if they have different pitch patterns, or "tones." In Chinese languages, the basic units that carry tone patterns are the monosyllabic morphemes. Each monosyllabic morpheme is represented orthographically by a "zi," that is, a character, which is written separately from its neighboring characters (in the following discussion, "morpheme" and "character" are used interchangeably depending on the context). Thus, the pronunciation of each character is defined not only in terms of consonants and vowels, but also in terms of the tone associated with it. Structurally, the orthography of Chinese characters is often composed of two parts, each with certain arrangement of strokes. One of them is often referred to as the "radical," which carries semantic information. The other parts is often referred to as the "phonetic," which carries phonological information. However, probably because of the changes accumulated over the years, information about pronunciation carried by the phonetic is anything but accurate. Moreover, no information about the tone of a character is directly carried by the phonetic. That information is available only after the recognition of the morphemic identity of the character.

There are a number of different languages and dialects in the Chinese language family, and each of them has a different tonal system. The language used in this study is Mandarin, which is the official language both in mainland China and in Taiwan. In Mandarin, there are four tones,[3] which are conventionally called Tone 1, Tone 2, Tone 3, and Tone 4. In "pinyin," a quasi-phonemic transcription system for Chinese characters, and also the transcription system used in this paper, the four tones are represented as marks over the nuclear vowels in the syllables. For example, *bā*, *bá*, *bǎ*, *bà* are the pinyin transcriptions of the syllable *ba* with Tone 1, 2, 3, and 4 respectively. Incidentally, those marks actually mimic the pitch contour of the four tones spoken in Mandarin.

In modern Chinese, the pronunciations of most of the morphemes are not unique. At the level of morphemes, homophones are commonplace. The number of different syllables with different tones recorded in *Cihai* (a grand dictionary of Chinese words; Xia, 1979) is 1,353, but the number of characters represented by those syllables in that dictionary is 14,872. Thus, an average syllable with tone represents about 11.02 different morphemes. To make things even worse, the information load of the syllables is far from being evenly distributed. For example, the syllable *néng* represents only one single morpheme, namely,

能

(to be able to, can). The syllable *yì*, on the other hand, represents 149 different morphemes, as recorded by *Cihai* (Xia, 1979). Fortunately, most of the Chinese words consist of more than one morpheme. Thus the number of homophones at the word level is much smaller than that at the morphemic level. However, the very fact that the Chinese orthography is structured at the morphemic level rather than at the word level, and that there are a large number of homophones at the morphemic level, makes some of our investigations easier or even possible.

Mandarin has a phonological rule called the "Third Tone Sandhi" rule. The rule says that whenever two Tone 3's are in succession within the same word and/or the same rhythmic unit, the first tone must change into a tone that is indistinguishable from Tone 2. Wang and Li (1967) found that Chinese listeners could not distinguish between a Tone 2 derived from Tone 3 by the tone sandhi rule and a nonderived morphemic Tone 2. They played, to Chinese subjects, pairs of bimorphemic words which were identical phonologically, except that in one of them the first morpheme is in Tone 2, whereas in the other, the first morpheme is in Tone 3, as in, for example, *máimǎ* (to bury a horse) and *mǎimǎ* (to buy a horse). The subjects' identification of those words was found to be very close to chance level.

The phonological ambiguity caused by application of the tone sandhi rule can be used in determining whether surface phonetic representation is generated in STM. If the surface representation is generated in STM, similarity at this surface level should adversely affect performance on immediate recall tasks. To verify this, certain linguistic material can be created in which the amount of tonal similarity varies between surface phonetic form and underlying morphophonemic form caused by the application of the sandhi rule. In other words, proper combinations of different tones can be manipulated so that there is more similarity in the material to be recalled after the application of the tone sandhi rule than there is before its application. If the surface phonetic representation is produced, then, immediate recall for this kind of material should be poorer than for material to which no such tone sandhi rule is applicable. If the rule is not applied, that is, if the surface phonetic representation is never produced, immediate recall should be the same for both kinds of material.

Before the test for tone sandhi is carried out, however, it is desirable to make sure that tonal similarity affects STM at all, which to this date has not been confirmed experimentally. Hence, two experiments are performed in this study. The first one investigates the effect of tonal similarity on STM; and the second one explores whether surface phonetic tonal similarity is produced through the application of the tone sandhi rule.

EXPERIMENT 1

Method

Material. Forty-eight test lists, each list consisting of seven Chinese characters and each character representing a monomorphic word, were used as stimuli in the experiment. In each list, all the characters shared the same vowel and final consonant, but differed in initial consonants. In other words, the characters in each list rhymed with one another. Two rhymes, *an* and *u*, were used in the lists. Those two rhymes were used because there are a sufficient number of characters of high and moderate frequencies with all four Mandarin tones that have the two rhymes, so that enough morphemes with different initial consonants could be found for each of the lists. The test lists were of two types. Lists of the first type were composed of characters with the same tone; lists of the second type were composed of the same characters used in the lists of the first type but in different combinations, so that each of the lists consisted of rhyming characters with different tones. Because there are only four different lexical tones in Mandarin, some of the characters in the second type of lists shared the same tone, but care was taken not to use any one tone more than twice in the same list.

Appendix A shows the sample lists that were used in Experiment 1. *List a* contains examples of the lists with the same tones, each of them consisting of characters with only one of the four lexical tones in Mandarin. The first four lists in *list a* all have the rhyme *an;* the last four rows all have the rhyme *u*. *List b* shows examples of the lists with different tones, each of them consisting of characters with different tones. As in *list a*, all the characters in the first four rows have the rhyme *an*, and those in the last four all have the rhyme *u*. The characters in *list b* are the same as those used in *list a* but in different combinations.

Subjects. In China, although everybody has a great amount of exposure to Mandarin[4] through radio and television, as well as varying amounts of training at school, most people speak their own language or dialect most of the time in their daily life until, in most of the cases, they enter college, where the common language is often Mandarin. Thus, it is not clear what kind of language or dialect they use when they read (assuming that they do use "inner language" of some sort in reading). As was mentioned in the introduction, each of the Chinese languages or dialects has its unique tonal system. Therefore, in a study involving tones, it is desirable not to mix subjects with different linguistic background, unless it is done deliberately.

Because the phonetic system of Mandarin is based on the native dialect spoken in Beijing, the speech of the people from Beijing, of all Chinese dialects, bears the closest resemblance to Mandarin. Thus, in any study on the pronunciation of Mandarin, native speakers of the Beijing dialect are the best subjects. Because this study was intended to test a hypothesis about the effect of phonological similarity on short-term memory, the ideal subjects were those who, not only in their conversation but also in their possible inner speech, speak the dialect whose phonetic system has the least deviation from the standard pronunciations that are recorded in the dictionary. Fortunately, there were a sufficient number of speakers of the Beijing dialect at the University of Connecticut for the purposes of this study. Twelve of them participated as subjects in Experiment 1. They were paid for their participation. Eight of them were females, and four were males. Most of them were graduate students at the University of Connecticut; the rest were college graduates. None of them reported any history of dyslexia.

Procedure. The method used in this experiment was similar to the paradigm introduced by Waugh and Norman (1965), which was used in many of the subsequent STM studies, including some of the studies on the STM of Chinese (Ren & Mattingly, 1990). The characters in each of the 48 lists were displayed on the computer screen one at a time. Each character stayed on the screen for 1 sec, and the following character appeared on the screen 0.5 sec after the disappearance of the previous character. One and a half seconds after the last character in the list disappears, two short beeps were heard; then, another 1.5 seconds later, a probe character appeared on the screen. This probe character was one of the characters that had occurred in the list. At this point, the subject was required to write down on the response sheet the character that followed the probe character in the list. Guessing was neither encouraged nor discouraged, and the subjects were allowed to leave a blank box on the answer sheet if they really could not remember the character. After each subject wrote down the character, the experimenter entered the pinyin transcription of that character into the computer. This input was later checked against the subjects' response sheet, and corrected accordingly if there was any disagreement. Only one probe was used for each of the lists, and because every list consisted of seven characters, there were six possible probe positions in each list.

With 48 trials, four judgments were obtained for every position under each of the two tonal conditions. Both the order of the appearance of the lists and the position of the probe characters were selected randomly by the computer program, and were different for different subjects.

The subjects were tested individually. Before the real test, each subject was given six practice trials to familiarize him or her with the procedure. At the very beginning of the test, the subject was given the following instructions in Chinese:

> This is a short-term memory test. During the test, Chinese characters will be shown on the screen one at a time. Please memorize every character as well as its position relative to other characters in the same group of characters. At the end of each group of characters (seven), you will hear two short beeps. Then, a reference character will appear on the screen. This is one of the characters in the group just shown. At this point, please recall which of the characters in the group just shown had occurred after this reference character and write it down.

Note that in the instructions, no specific requirements were made as to whether the subject should or should not pronounce the characters while trying to remember them. This was to allow the subject to do the task in a way they felt was most natural to themselves. As it turned out, most of them did not pronounce the characters, although some of them did seem to be moving their lips during the test as if trying to murmur to themselves.

Results

The subjects' responses with characters that were different from the stimulus characters but had pronunciations, including tones, identical to those of the stimulus characters, were scored as correct responses. Figure 1 shows the percentage of correct responses against the serial order for the recall test. Unlike most STM studies, which typically show a serial-position effect, the pooled result here did not show an apparent serial effect. An ANOVA analysis showed that the main effect of serial positions was not significant, and that there was no significant interaction between serial position and tonal similarity. Several factors might have contributed to this result. First, for each of the six serial positions in each of the two categories, there were only four trials. This was probably not enough to reveal a strong serial effect. Second, the beeps used before the presentation of the probe character could account for the absence of the normal rise in the last two positions. Some of the subjects reported that the beeps seemed to have wiped out their memory.

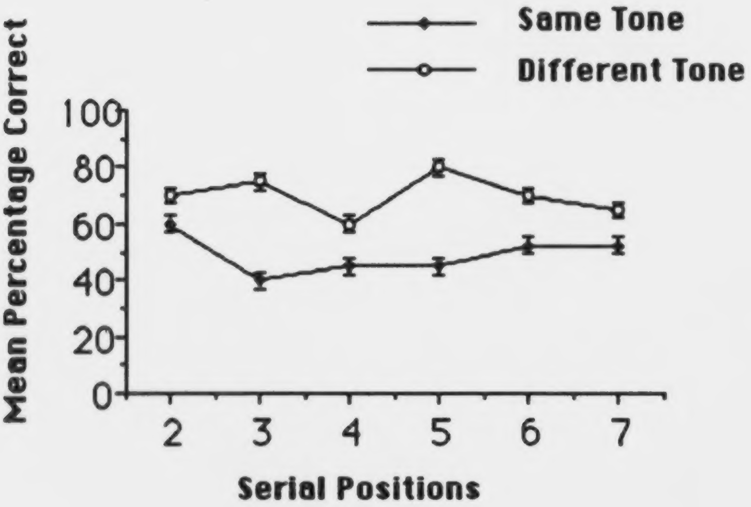


Figure 1. Mean percentage of correct responses for stimulus lists with mono-tonal and multi-tonal lists plotted against serial positions. (Horizontal bars represent standard error.)

Third, subjects seemed to have used different recall strategies when trying to remember the lists. Some of them reported that they had grouped the first four characters in the lists into one set and the last three characters into another set; others reported that they did the opposite; still others reported that they had tried to remember only the pronunciations of the first six characters, leaving the last one to visual memory. This could have resulted in different patterns of recall as a function of serial positions of the items in the lists.

The lack of a significant main effect of order did not affect the main effect of tonal similarity. As is obvious from Figure 1, the tonal similarity has significant effect on the recall performance, $F(1, 11) = 25.28$, $p < 0.001$. When asked to comment on their experience during the experiments, all of the subjects reported that the task was very hard, because the lists were very difficult to say, although only some of them had made apparent effort to pronounce the lists quietly. Interestingly, none of the subjects realized that some of the lists were monotonal whereas the others had different tones, and few of the subjects noticed that all the lists rhymed internally. This is probably evidence that conscious phonological observation might not be directly related to the actual coding strategy used in reading.

Discussion

The results of experiment 1 indicate that tone, like rhyme, plays an important role in STM for Chinese. This conclusion is not very surprising because, in a tone language, the tonal features are part of the phonological specifications that enabled one to distinguish between different morphemes. Baddeley (1966) found that vowel similarity in the word sequence to be recalled significantly affects the recall performance. Brady et al. (1987) reported that the number of errors in ordered recall of auditorially presented syllables is a function of the phonetic similarity of consonantal features. The effect of tonal similarity on the STM of linguistic material in a tone language is simply a further demonstration that various aspects of the phonetic system, as long as it is relevant to speech, are likely to be involved in the verbal short-term memory.

Because Experiment 1 confirmed the effect of tonal similarity on the STM of the speakers of a tone language, this finding could now be exploited to test the hypothesis that the Mandarin tone sandhi rule is applied in STM by checking if the effect of tonal similarity shows up again, this time as a result of a special design of the tests in Experiment 2.

EXPERIMENT 2

Material. The test items used in this experiment were 48 lists, each consisting of 3 disyllabic nonsense words. Pilot tests had shown that lists with 4 disyllabic nonsense words were a bit too difficult for the subjects to recall. This agrees with the finding of Zhang and Simon (1985) that STM span for two-character Chinese words was 3.83 whereas for isolated characters it was 6.38. In their study, the 2-character words were real words, whereas here they are nonwords, which should reduce STM span even further. Three 2-character nonwords was found to be an appropriate size for this experiment.

Appendix B shows all the test lists used in this experiment. The lists were designed so that most of the errors would be confusions on the tones of the first characters in the nonsense words. Each element of a list consists of two characters that, although meaningful in themselves, do not form a meaningful bimorphemic word. For example, the characters in column 1 of the first row in Appendix B mean, respectively, "frequent" and "willing to." This combination is not a meaningful word in Chinese. In each list, the first characters in all three words have the same syllable structure (i.e., the same rhyme, as well as the same initial consonant) except for their tones, which are either Tone 2 or Tone 3. The graphic shapes of the characters have no similarity to one another. The second characters in each list all share the same tone but have completely different syllable structures, that is, they share neither the same initial consonant, nor the same rhyme among themselves or with the first characters that precede them. Two tones are used for the second syllables in the lists, depending on the classification of the test lists. In the first type of lists (Sandhi List, Appendix B), the second syllables all have Tone 3. This means that, in these lists, whenever the first character is in Tone 3, the possibility exists that the tone sandhi rule might apply as it always does in speech. If this happens, the first character should become indistinguishable phonetically from the other first character(s) in the same list that have an underlying Tone 2. In this way, the first three characters in each list of this type would be phonetically identical to one another, resulting in possible confusion among them. In the second type of list (No-Sandhi List, Appendix B), the second characters all have Tone 4. In this case, no tone sandhi rule is applicable. Thus, the first characters in those lists will remain different in tone whether or not the tone sandhi rule applies in STM. As can be noticed, the first characters in the No-Sandhi Lists and the Sandhi Lists are exactly the same ones, but the second characters are all different. Thus, the only real difference between the two types of lists is in the tones of the second characters, which make the tone sandhi either applicable (in the Sandhi Lists) or non-applicable (in the No-Sandhi Lists).

Subjects. Ten Chinese students, three male and seven female, participated in Experiment 2. Again, they were native speakers of the Beijing dialect. They were also paid for their participation.

Procedure. The procedure used in Experiment 2 was similar to that used in Experiment 1. Each of the disyllabic nonsense words was displayed on the computer screen for one sec. After a delay of .5 sec, the following word appeared on the screen. One and a half seconds after the disappearance of the last word in a list, two short beeps were heard and, 1 sec later, a probe word appeared on the screen and stayed there for 2 sec. The subjects' task was to write down the word that followed the probe word in the list just shown. When the probe item was "• •," the subject was to write down the first word in the list. This was to guarantee that the subject would try to remember all the three words in the list rather than just the last two. This was necessary because, when there are only three words in the list, it is possible to ignore the first one and still accomplish the task. When the probe item is recognized as a strange word, the subject can safely assume it to be the first one in the list. This tactic was actually used, and rather successfully, by quite a few subjects in a pilot study, thus creating a confounding subcategory among the subjects.

The subjects were tested individually. Each subject was given six practice trials before the start of the real test. At the beginning of the experiment, written instructions in Chinese similar to those used in Experiment 1 were given to the subject. As in Experiment 1, no requirements were made as to whether subject should or should not pronounce the words, thus allowing the subject to do the task in a natural way. Some of the subjects did move their lips quietly during the test; others remained totally silent. After the subject wrote down each response, the experimenter entered it into the computer. This input was later checked against the subjects' response sheets and corrected accordingly. As in Experiment 1, guessing was not discouraged, but neither was it encouraged, and the subjects were allowed to leave a blank if they really did not think they could remember the word.

Results

The subjects' responses were scored in three different ways: First, any disagreement between the pronunciation of the response character and the stimulus character was counted as an error; second, only disagreements on the first characters were counted as errors; and third, only disagreements on tones of the first characters were counted as errors. The results of two-factor within-subject ANOVA for the data scored with those three methods were all very close. The following discussion will only use data obtained with the second method, that is, treating any disagreement on the first character as an error. Figure 2 shows the mean percent correct responses plotted against serial positions for the Sandhi Lists and No-Sandhi Lists. Again, the effect of serial positions on STM was not significant. This time, the lack of a significant serial effect might be due to list length, which was not long enough to create clear contrasts among different serial positions. And, again, the beeps after the presentation of the word lists might have contributed to the absence of the final rising in the recall curve. The interaction between serial position and list type also was not significant.

There is a significant main effect of list type, $F(1, 9) = 20.02$, $p < 0.01$. The number of errors on the first characters in the Sandhi Lists was two and half times the number of errors on the first characters for the No-Sandhi Lists. Among all the errors on Sandhi Lists, 77.2% were those on the tones of the first characters, whereas only 54.24% of the errors on the No-Sandhi Lists were those on the tones of the first characters.

Almost all the errors on the tones of the first character were confusions between Tone 2 and Tone 3. This is true for both Sandhi Lists and No-Sandhi Lists. Of the total of 73 errors on tones of the first characters in Sandhi Lists, 43 were substitution of Tone 2 for Tone 3, and 30 were substitution of Tone 3 for Tone 2. A one factor repeated measure ANOVA showed no significant difference between the two, indicating that the confusions were in both directions.

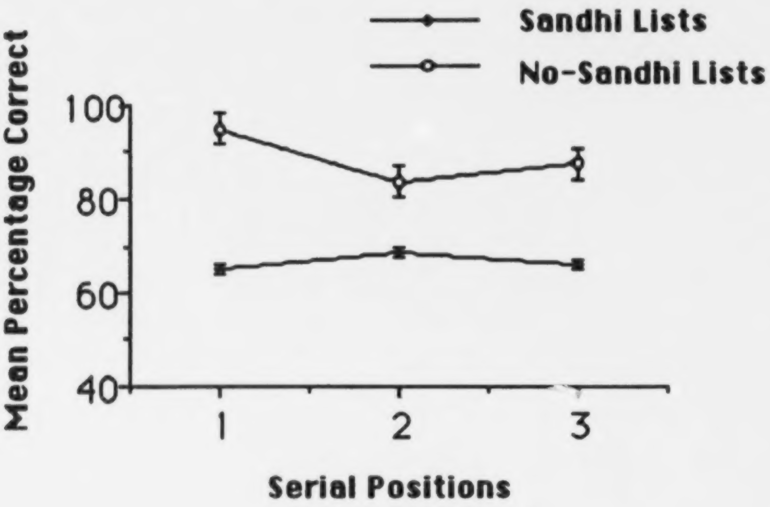


Figure 2. **Mean percentage correct responses plotted against serial positions for Sandhi lists and No-Sandhi lists. (Horizontal bars represent standard error.)**

Discussion

Logically, to fulfill this particular immediate recall task, the best thing to remember would have been the visual forms or the semantic values of the characters, because they were all different from one another, so there was little possibility of confusion. The second best thing to remember would have been the underlying phonological forms, because there were at least two different tones on the first characters in the Sandhi Lists. But the subjects seemed to have used the worst thing, that is, the surface phonetic form, which was identical within each Sandhi Lists for the first characters. This can only imply that in performing the immediate recall task, the derivation of the surface phonetic forms through the application of the phonological rules is somehow compulsory.

When asked to comment on the experiment, most of the subjects reported that some of the lists seemed easier to remember because the first characters in those lists all seemed to be the same, whereas in some other lists, the first characters seemed to have different tones, so that the subject had to pay special attention to remember the tonal differences. Apparently, the three first syllables in each of the Sandhi Lists were taken by the subjects as having the same tone as well as having the same segmental structure. Two things here are of interest to us. First, by reducing the difference between the pronunciations of the characters to only two tones, speakers of a language with a non-alphabetic writing system seemed to have been forced to come to awareness about some of the phonological features in their language. Second, the tone sandhi rule is so natural, even in a process that does not require overt phonetic production, that although all of the subjects applied it, they were nevertheless totally unaware of its occurrence.

GENERAL DISCUSSION

The two experiments performed in this study reveal a strong effect of tonal similarity on STM for visually presented material in a tone language. This effect is found to be present both when the surface phonetic form and the underlying morphophonemic form agree and when the two do not agree. In the latter case, the similarity in surface phonetic form affected STM in the absence of the same similarity in underlying morphophonemic form.

The results of this study further confirm the finding of previous studies that sensitivity of STM to phonological similarity is not a function of the nature of the writing system for the language being tested (Erickson et al., 1977; Hue & Erickson, 1988; Ren & Mattingly, 1990; Tzeng et al., 1977; Yik, 1978; Zhang & Simon, 1985). The findings of this study also add a new aspect of the phonology, namely, tonal features in tone languages, to the list of the feature classes that STM is sensitive to. Furthermore, besides investigation of the effect of phonological similarity in general, the question about the levels of the phonology involved in STM is also addressed in this study. The results of Experiment 2 demonstrate that surface phonetic representation is generated in STM for visually presented material and codes in surface phonetic form are actually used in STM.

There are two aspects of the findings in this study that are worth noticing. First, the tonal features of Chinese are, compared with other aspects of its phonology, even less directly represented in the orthography for this language, for even the phonetic radical of a character does not give any reliable clue about the tone of the morpheme represented by the character. Nevertheless, the tonal features of the characters were apparently activated during the tests, because the tonal similarity in the test material severely impaired the immediate recall of the material. Second, not only were the tones of the isolated characters activated, but also the tone sandhi rule was apparently applied, because the tonal similarity that is possible only after the tone sandhi also seriously impaired the recall of the material. Two questions could be asked. First, why should aspects of the phonology that have no direct representation in the orthography be activated at all? Second, why should the readers or recallers go all the way to generate the surface phonetic representation rather than stick to the underlying morphophonemic representation which, in that particular test situation, actually provides more relevant information for distinguishing between the test items? Evidently, in performing the STM task, the subjects were resorting to a phonological mechanism similar to the one used in speech production and speech perception. In other words, it seems that so-called phonological encoding involves not only the underlying abstract phonology of a language but also the surface phonetic representation which includes, though not necessarily in real acoustic forms, all the relevant features crucial to speech production and perception.

One accounts for the phonological nature of STM is the one proposed by Baddeley and Hitch (1974). They described the phonological mechanism as a buffer in the working memory. Controlled by the central executive component of the working memory, the phonological mechanism is said to be responsible for storing material for a few seconds while awaiting higher-level processing. A more recent attempt to explain the nature of the phonological mechanism in STM has been made by Ren and Mattingly (1990). They argued that the phonological mechanism is a consequence of the operation of the language module. According to their account, so-called short-term memory is simply a process for exploiting the language module, which is used both in the primary process of speaking and listening and in the secondary process of reading and writing. As shown by both current research and previous research, in situations where visually presented linguistic material is to be remembered for more than a few milliseconds, the language module is also employed, which is perhaps the only choice, considering the very limited capacity of the short-term iconic storage for temporally presented information (Averbach & Coriell, 1961; Frick, 1985; Zhang & Simon, 1985). The nature of the language module is such that the computation and output of the various representations is automatic and compulsory (Fodor, 1983, Liberman & Mattingly, 1985; Mattingly & Liberman, 1990). In this manner, then, the output of the surface phonetic form is inevitable. Even in situations that would favor, at least logically, the suppression of the generation of the surface form, it is nevertheless generated at the cost of causing phonological confusion.

Although the findings of this study provide us with more evidence for the phonological nature of STM, especially for the inclusion of surface phonetic representation in STM, there are still important details about the process of STM that we do not yet understand well. For example, the modular account predicts the generation of both the underlying representation and the surface representation of linguistic material in STM. If this is true, would properties of the underlying representation also affect STM? For instance, would the span of STM be a function of the similarity in the underlying forms when the similarity in surface forms is held constant? Or, would the number of phonemes affect STM span when the surface duration of words are held constant? Finding answers to these questions would certainly further our knowledge of the phonological nature of short-term memory.

Finally, the finding that certain phonological rules are applied during the encoding of visually presented material by readers of a language with a nonalphabetic writing system also provides further evidence for the similarity of reading processes between languages with different writing systems. It is easy to believe that readers of languages with alphabetic writing systems access the pronunciation of words analytically, because the writing system more or less directly reflects the phonological structures of words. For a language with nonalphabetic writing system, on the other hand, it is easy to believe that a "whole word" approach should be used; that is, the pronunciation is only activated through lexical access. However, there is evidence that even for languages with alphabetic writing system, pronunciations of familiar words are activated through lexical access rather than through applying pronunciation rules to the spelling of the words (Baddeley & Lewis, 1981; Besner, Davies, & Daniels, 1981). To pronounce unfamiliar words, even for Chinese, on the other hand, readers have to resort to the phonetic component (if there is one) of the character to get the pronunciation (Seidenberg, 1985; Shu & Zhang, 1987), although frequently the guessed pronunciation is not the correct one. The finding of this study, especially that of Experiment 2, further suggests that even for a language like Chinese, the phonological structures at the word level could be analytical, although, perhaps, not cognitively analytical. The pronunciation of a sequence of monosyllabic morphemes is not necessarily a simple combination of the pronunciations of the component morphemes; rules have to be applied to derive the surface phonetic structure. Thus, rather than always being accessed directly, the pronunciation of the words may be derived through the application of pronounciation rules, especially for unfamiliar words. This is probably similar to the stress assignment in languages such as English. In this sense, there should be no basic difference in the process of lexical access for languages with a nonalphabetic writing system, such as Chinese, and for languages with an alphabetic writing system, such as English.

REFERENCES

Averbach, E., & Coriell, A. (1961). Short-term memory in vision. *Bell System Technical Journal, 40,* 309-328.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic and formal similarity. *Quarterly Journal of Psychology, 18,* 362- 365.

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology.* 20, 249- 264.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

Baddeley, A. D., & Hitch, G. B. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and activation, 4* (pp. 47-90). New York: Academic Press.

Baddeley, A. D., & Lewis, V. J. (1981). Inner active processes in reading: The inner voice, the inner ear and the inner eye. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 107-129). Hillsdale, NJ: Erlbaum.

Besner, D., Davies, J., & Daniels, S. (1981). Phonological processes in reading: The effects of concurrent articulation. *Quarterly Journal of Experimental Psychology, 33,* 415-438.

Brady, S., Mann, V., & Schmidt, R. (1987). Errors in short-term memory for good and poor readers. *Memory & Cognition, 15*(5), 444-453.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75-84.

Conrad, R. (1972). Speech and reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading* (pp. 205-240). Cambridge, MA: MIT Press.

Erickson, D., Mattingly, I. G., & Turvey, M. (1977). Phonetic activity in reading: An experiment with Kanji. *Language and Speech, 20,* 384- 403.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Frick, R. W. (1985). Testing visual short-term memory: Simultaneous versus sequential presentations. *Memory & Cognition, 13(4),* 346-356.

Hue, C-W., & Erickson, J. R. (1988). Short-term memory for Chinese characters and radicals. *Memory & Cognition,16(3),* 196-205.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Locke, J. L., & Kutz, K. J. (1975). Memory for speech and speech for memory. *Journal of Speech and Hearing Research, 18,* 176-191.

Mattingly, I. G., & Liberman, A. M. (1990). Speech and other auditory modules. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps* (pp. 501-520). New York: Wiley.

Ren, N., & Mattingly, I. G. (1990). Short-term serial recall performance by good and poor readers of Chinese. *Haskins Laboratories Status Report on Speech Research, SR-103/104, 153-164.*

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19,* 1-30.

Shankweiler D., Liberman I. Y., Mark, L. S., Fowler, C. A., & Fisher, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 531- 545.

Shu, H., & Zhang, H-C. (1987). The process of pronouncing Chinese characters by proficient mature readers. *Acta Psychologica Sinica, 19(3),* 282-290.

Tzeng, O. J. L., Hung, D. L., & Wang, W. S-Y. (1977). Speech recoding in reading Chinese characters. *Journal of Experimental Psychology: Human Learning and Memory,* 3, 621-630.

Wang, W. S-Y & Li, K-P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research, 10,* 629-636.

Waugh, H. C., & Norman, D. H. (1965). Primary memory. *Psychological Review, 72,* 89-104.

Wickelgren, W. A. (1965a). Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America, 38,* 583- 588.

Wickelgren, W. A. (1965b). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America, 39,* 388-398.

Xia, Z-N (Ed.) (1979). *Cihai [*A grand dictionary of Chinese words*]*. Shanghai: Shanghai Dictionary Press.

Yik, W. F. (1978). The effect of visual and acoustic similarity on short-term memory for Chinese words. *Quarterly Journal of Experimental Psychology, 30,* 487-494.

Zhang, G. J., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition, 13(3),* 193-201.

FOOTNOTES

[*] Appears in *Memory & Cognition, 19*(3), 263-273 (1991).

[†] Also Department of Linguistics, University of Connecticut, Storrs.

[1] The significant effect taken to be that of visual similarity in Yik (1978) might have been confounded by some other factors. Of the five pairs of monosyllabic words considered to contain high visual similarity and low acoustic similarity, four pairs have either the same consonant (e.g., *shuai* vs. *shi*), or the same glide (e.g., *xuan* vs. *lu*), and one pair is semantically similar (*fen*: powder vs. *li:* small particle). In the inventory with low visual and high acoustic similarity, which turned out to have produced fewer errors than the inventory with high visual and high acoustic similarity, two pairs are not acoustically identical (e.g., *zhi* vs. *zi*).

[2] Incidentally, their homophone set contains a character that has alternative tones, i.e., *gōng* and *gòng*.

[3] At the word level, there is a fifth tone which is usually called the neutral tone. This tone occurs only after a full lexical tone within a word.

[4] Mandarin, or Putonghua (the Common Chinese), as it is called in mainland China, or Guoyu (the National Language), as it is called in Taiwan, is an official language based on the Beijing dialect and other northern dialects of China. It is used mainly on radio, TV and on other public occasions.

APPENDIX A

Sample Lists of Monosyllabic Morphemes (a) with Same Tones and (b) with Different Tones

(a)

1)	班 bān	肝 gān	安 ān	贪 tān	山 shān	翻 fān	三 sān
2)	残 cán	谈 tán	篮 lán	然 rán	凡 fán	含 hán	男 nán
3)	产 chǎn	胆 dǎn	砍 kǎn	伞 sǎn	反 fǎn	敢 gǎn	展 zhǎn
4)	烂 làn	饭 fàn	慢 màn	叛 pàn	站 zhàn	岸 àn	汉 hàn
5)	粗 cū	姑 gū	屋 wū	租 zū	书 shū	出 chū	突 tū
6)	读 dú	胡 hú	奴 nú	图 tú	如 rú	除 chú	吴 wú
7)	府 fǔ	苦 kǔ	母 mǔ	主 zhǔ	普 pǔ	鲁 lǔ	补 bǔ
8)	布 bù	树 shù	路 lù	度 dù	促 cù	护 hù	速 sù

(b)

1)	肝 gān	残 cán	展 zhǎn	山 shān	慢 màn	然 rán	反 fǎn
2)	产 chǎn	凡 fán	叛 pàn	班 bān	含 hán	岸 àn	贪 tān
3)	汉 hàn	谈 tán	砍 kǎn	三 sān	敢 gǎn	翻 fān	烂 làn
4)	男 nán	站 zhàn	胆 dǎn	安 ān	饭 fàn	篮 lán	伞 sǎn
5)	粗 cū	度 dù	吴 wú	补 bǔ	如 rú	书 shū	普 pǔ
6)	主 zhǔ	树 shù	护 hù	除 chú	突 tū	租 zū	读 dú
7)	路 lù	胡 hú	出 chū	速 sù	屋 wū	府 fǔ	苦 kǔ
8)	奴 nú	图 tú	促 cù	姑 gū	鲁 lǔ	母 mǔ	布 bù

APPENDIX B
Test Lists for Experiment 2

Sandhi List

1.	常肯 chángkĕn	肠者 chángzhĕ	厂把 chăngbă
2.	迟永 chíyŏng	齿法 chĭfă	池给 chígĕi
3.	楚品 chŭpĭn	除甲 chújiă	厨本 chúbĕn
4.	宇也 yŭyĕ	语广 yŭguăng	余久 yújiŭ
5.	府伙 fŭhuŏ	浮北 fúbĕi	抚瓦 fŭwă
6.	级果 jíguŏ	几可 jĭkĕ	挤凯 jĭkăi
7.	节免 jiémiăn	杰挺 jiétĭng	姐否 jiĕfŏu
8.	兰纽 lánniŭ	览整 lănzhĕng	蓝引 lányĭn
9.	礼丙 lĭbĭng	厘走 lízŏu	梨写 líxiĕ
10.	止小 zhĭxiăo	纸吕 zhĭlŭ	执景 zhíjĭng
11.	每拐 mĕiguăi	眉考 méikăo	美总 mĕizŏng
12.	其忍 qírĕn	启朵 qĭduŏ	企铁 qĭtiĕ
13.	逐举 zhújŭ	竹远 zhúyuăn	煮冷 zhŭlĕng
14.	石怎 shízĕn	史顶 shĭdĭng	实很 shíhĕn
15.	统角 tŏngjiăo	童两 tóngliăng	铜左 tóngzuŏ
16.	碗准 wănzhŭn	晚轨 wănguĭ	完表 wánbiăo
17.	网吵 wăngchăo	亡妥 wángtuŏ	往短 wăngduăn
18.	维选 wéixuăn	伪损 wĕisŭn	尾且 wĕiqiĕ
19.	无软 wúruăn	吴捆 wúkŭn	午纽 wŭniŭ
20.	习缓 xíhuăn	洗逞 xĭchĕng	席检 xíjiăn
21.	显耳 xiănĕr	闲孔 xiánkŏng	咸仅 xiánjĭn
22.	演所 yănsuŏ	掩响 yănxiăng	严口 yánkŏu
23.	以扫 yĭsăo	移暖 yínuăn	椅柳 yĭliŭ
24.	尤审 yóushĕn	友粉 yŏufĕn	有董 yŏudŏng

No-Sandhi List

1.	常并 chángbìng	肠日 chángrì	厂寸 chǎngcùn
2.	迟后 chíhòu	齿过 chǐguò	池页 chíyè
3.	楚问 chǔwèn	除令 chúlìng	厨运 chúyùn
4.	宇或 yǔhuò	语克 yǔkè	余动 yúdòng
5.	府向 fǔxiàng	浮备 fúbèi	抚困 fǔkùn
6.	级认 jírèn	几正 jǐzhèng	挤定 jǐdìng
7.	节最 jiézuì	杰命 jiémìng	姐众 jiězhòng
8.	兰奋 lánfèn	览彻 lǎnchè	蓝次 láncì
9.	礼况 lǐkuàng	厘序 líxù	梨这 lízhè
10.	止共 zhǐgòng	纸列 zhǐliè	执外 zhíwài
11.	每壮 měizhuàng	眉册 méicè	美劝 měiquàn
12.	其会 qíhuì	启化 qǐhuà	企罢 qǐbà
13.	逐竟 zhújìng	竹幸 zhúxìng	煮样 zhǔyàng
14.	石对 shíduì	史面 shǐmiàn	实律 shílǜ
15.	统近 tǒngjìn	童够 tónggòu	铜念 tóngnìan
16.	碗甚 wǎnshèn	晚个 wǎngè	完落 wánluò
17.	网要 wǎngyào	亡料 wángliào	往碰 wǎngpèng
18.	维秀 wéixiù	伪顿 wěidùn	尾用 wěiyòng
19.	无下 wúxià	吴透 wútòu	午欠 wǔqiàn
20.	习大 xídà	洗宋 xǐsòng	席去 xíqù
21.	显断 xiǎnduàn	闲旧 xiánjiù	咸社 xiánshè
22.	演庙 yǎnmiào	掩庆 yǎnqìng	严创 yánchuàng
23.	以贵 yǐguì	移配 yípèi	椅灭 yǐmiè
24.	尤论 yóulùn	友扣 yǒukòu	有赤 yǒuchì

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 209-230

Phonological Access of the Lexicon: Evidence from Associative Priming with Pseudohomophones*

Georgije Lukatela† and M. T. Turvey††

Four experiments were conducted with pseudohomophones in a primed naming task. In Experiments 1 and 2, target pseudowords that sounded like real words, for example, CHARE, were preceded either by context words that related associatively to the word with which the target was homophonic, TABLE-CHARE, or by context words that were not associatively related, NOVEL-CHARE. Control pairs were TABLE-THARE and NOVEL-THARE (Experiment 1) and TABLE-CHARK and NOVEL-CHARK (Experiment 2). The prior presentation of TABLE relative to the prior presentation of NOVEL benefited the naming of CHARE but not the naming of THARE or CHARK. The third experiment placed pseudohomophones in the role of primes, that is, TAYBLE-CHAIR, with such pairs comprising only 8% of all pairs seen by a subject in order to counter guessing strategies. If the prime TAYBLE activated /table/, then /chair/ would be activated associatively and the target CHAIR would be named faster than if TARBLE was the prime. This result was obtained. The fourth experiment extended the design of the third to include TABLE-CHAIR pairs and a comparison of a short (280 ms) and a long (500 ms) delay between context and target onsets. The priming due to associated pseudohomophones was unaffected by onset asynchrony and equal in magnitude to that due to associated words. The overall pattern of results suggests that lexical representations are coded phonologically and accessed phonologically.

Most research on recognizing and pronouncing words has been motivated by the dual process theory (Coltheart, 1978). According to this theory, two independent processes—a direct, visual process and a mediated, phonological process—govern the accessing of the internal lexicon, a mental dictionary containing relatively permanent information about the identity of individual words. The two independent processes are not used equally. The direct process which entails a mapping between orthographic features of individual printed/written words and lexical representations is the primary access route. The secondary access route is provided by the mediated process which involves a set of grapheme-phoneme correspondence rules turning spellings into phonological representations, and a subsequent mapping from these phonological representations onto lexical entries. According to the theory, whereas the mediated route might dominate word identification in early stages of reading, it is the direct route which characterizes reading fluency. And whereas phonological mediation is needed for reading new words and nonwords, the direct visual route is mandatory for exceptional spellings and is preferred for familiar words.

A recent major review of the literature concluded that the dual process theory is essentially wrong (Humphreys & Evett, 1985). The heart of the criticism was that the appeal to a rule-governed phonological process to access the lexicon is superfluous. Lexical access is achieved for all word forms in a word-specific manner, by the direct visual route. Evidence from studies with Serbo-Croatian language materials, and new

---

This research was supported in part by National Institute of Child Health and Human Development Grants HD-08945 and HD-01994 to the first author and Haskins Laboratories respectively. The authors wish to acknowledge Arthur Abramson's help in constructing the pseudohomophones used in Experiment 1.

evidence from studies with English language materials suggest, however, a different criticism of the dual process theory: The primary constraint on lexical access is phonological, not visual, and this constraint arises not through explicit rules but through continuous statistical regularity (Van Orden, Pennington, & Stone, 1990).

Until recently empirical support for phonological mediation was sparse. The early demonstration by Rubenstein, Lewis, and Rubenstein (1971) that lexical decisions are slower for homophones (e.g., YOLK) and pseudohomophones, that is, nonwords homophonous with real words (e.g., TRATE), was originally taken as evidence for speech-related processes in word identification. Subsequent research, however, found that the homophony effect on "yes" responses could be eliminated by including many pseudohomophones as foils (Davelaar, Coltheart, Besner, & Jonasson, 1978) suggesting that the phonological route was optional, and probably used only for the processing of nonwords. The prelexical phonological processing of a nonword such as TRATE produces a representation which activates the lexical entry for the word TRAIT, making the rejection of TRATE difficult. Unfortunately, because this evidence for prelexical phonology is provided by the slower rejection latencies, it gives rise to doubts about phonology's role in actual word identification. Various investigators have argued that even if phonological coding does occur, it occurs too slowly to be of use in lexical access (e.g., Henderson, 1982; McCusker, Hillinger, & Bias, 1981). As Coltheart, Davelaar, Jonasson, and Besner (1977, p. 551) remarked: "Unequivocal evidence for this view would be obtained by demonstrating that the phonological code for a word is sometimes used in making the "yes" response to that word in a lexical decision or categorization task; such a demonstration remains to be achieved."

Recent experiments on rapid semantic categorization have provided such a demonstration (Van Orden, 1987; Van Orden, Johnston, & Hale, 1988). The basic observation is that subjects produce larger false positive error rates when they respond to foils that are homophonic to category exemplars (e.g., ROWS for the category A FLOWER) than when they respond to spelling control foils (e.g., ROBS). Further, false positive errors to nonword homophone foils (e.g., SUTE for AN ARTICLE OF CLOTHING) exceed false positive errors to nonhomophonic nonword spelling controls; the phonological characteristics of the nonword foils are critical. Many other demonstrations of phonological influences on positive responses are to be found in lexical decision and naming experiments with Serbo-Croatian materials.

Because of the use of two, partially overlapping alphabets, and because of the one letter-one phoneme principle for both alphabets, it is possible to construct letter strings in Serbo-Croatian that can be read legally in more than one way. Take BETAP as an example. Read strictly through the letter-to-sound correspondences of the Cyrillic alphabet, this letter string is pronounced /vetar/ and is a high frequency word meaning "wind.". Read strictly through the letter-to-sound correspondences of the Roman alphabet, BETAP is pronounced /betap/, a nonword. Read with a mixture of the two sets of correspondences, Cyrillic and Roman, leads to the pronunciations /vetap/ and /betar/, which are also nonwords. The word meaning "wind" is transcribed in the Roman alphabet as VETAR. This letter string supports only a single reading, /vetar/. No other readings are possible. VETAR, unlike its Cyrillic mate BETAP, is phonologically unambiguous. In the lexical decision and naming tasks, the latencies for BETAP, and for letter strings like it, are considerably longer than the latencies for VETAR, and for letter strings like it, even though the two letter strings are equal in frequency, syllabic structure, number of letters, and meaning. This contrast has been called the phonological ambiguity effect.

The effect of phonological ambiguity on "yes" responses in lexical decision and naming implies that phonology mediates word identification. The same conclusion holds for nonwords. Rejection latencies are slowed for the phonologically ambiguous nonword BEMAP relative to its phonologically unambiguous nonword mate VEMAR. (Arguably, BEMAP is coded into more phonological forms than VEMAR and, because of this, BEMAP's rejection latencies are slowed relative to those for VEMAR.) The fact that the phonological ambiguity effect occurs for both words and nonwords, suggests that phonological mediation must be routine. Significantly, the phonological ambiguity effect is generally larger for words than nonwords (e.g., Feldman & Turvey, 1983; Lukatela et al., 1980; Lukatela, Feldman et al., 1989). As noted above, experiments showing that homophony influences nonword processing more so than word processing have been interpreted to mean that phonological coding must be rare in ordinary word recognition (Coltheart et al., 1977, Henderson, 1982; McCusker et al., 1981). Applying this logic to the Serbo-Croatian results, phonological coding must be routine in ordinary word recognition because phonological ambiguity influences the processing of words more than the processing of nonwords.

It has often been suggested that phonological mediation should reveal effects dependent on the number of letters or syllables in a word (Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Green & Shallice, 1976). Letter length effects have been observed regularly in naming but hardly ever in lexical decision experiments using English language materials (Henderson, 1982). The lack of such effects in lexical decision is used to argue that, contrary to dual process theory, there is no phonological route to the lexicon. Results from research with Serbo-Croatian materials demonstrate "number-of-constituents effects" and lead, thereby, to the contrary conclusion. KOTBA has one ambiguous letter (B), BETAP has two ambiguous letters (B and P), and CABAHA has three (C, B, and H). In each example the remaining unambiguous letters are shared letters. In the lexical decision task with such stimuli, the phonological ambiguity effect is larger the greater the number of ambiguous letters (Feldman & Turvey, 1983; Feldman, Kostić, Lukatela, & Turvey, 1983). The implication is that there is a prelexical process that is phonologically analytic (Turvey, Feldman, & Lukatela, 1984).

Other evidence points in the same direction. We would expect a phonological route to the lexicon to be largely indifferent to variables related to words, such as grammatical role and familiarity. The inflected nouns of Serbo-Croatian are most frequent and most prominent grammatically in the nominative singular form. A common finding is that nominative singulars are responded to faster than oblique forms (e.g., Lukatela, Carello, & Turvey, 1987). For example, VENA ("vein" in nominative singular) is responded to much faster than VENI (same word in dative or locative singular). However, if VENA and VENI are written in Cyrillic, that is, BEHA and BEHИ, respectively, then the less frequent and less grammatically prominent dative/locative form BEHИ is responded to faster. The latency difference between BEHA and VENA is large (the phonological ambiguity effect); in contrast, the latencies to BEHИ and VENI do not differ (Feldman et al., 1983). The process underlying the phonological ambiguity effect is not affected by the grammatical significance of word stimuli. Further, the effect has been shown to be indifferent to frequency; it is of the same magnitude for both frequent and infrequent words (Lukatela & Turvey, 1987).

There are two kinds of error patterns that enforce the notion of phonological mediation. Rejecting a BETAP-type word as a word occurs on a high proportion of the times such a word is presented, for example, 20% (Lukatela et al., 1978, Experiment 1), 19% (Lukatela et al., 1978, Experiment 2), 26% ((Lukatela, Popadić, Ognjenović, & Turvey, 1980), 22% (Lukatela, Feldman, Turvey, Carello, & Katz, 1989, Experiment 1), and 21% (Lukatela, Feldman et al., 1989, Experiment 2). In these examples, average false negatives on phonologically unambiguous words were in the range of 1% to 4%. False negatives of the order of 20-25% suggest a process in which the representation activating the lexicon frequently does not correspond to a word. This is understandable if (a) BETAP can give rise to a number of phonological representations that do not correspond to a word, and (b) the lexicon is accessed through phonological representations. The same conclusion follows from a consideration of false positive responses.

There are two significant types of phonologically ambiguous nonwords in Serbo-Croatian. One type comprises letter strings that are (a) composed of shared letters, one or more of which are ambiguous, (b) unreadable as a word in either Roman or Cyrillic, and (c) unreadable as a word in a mixture of both alphabets. The other type satisfies (a) and (b) but not (c). Consider the nonword HAPEB as an example of the second type. If the letter-phoneme correspondences of both alphabets are applied to HAPEB, then one of the resulting six phonological descriptions corresponds to a word, namely, /napev/ meaning "tune." It comes about by assigning the phoneme /n/ to H by the Cyrillic alphabet, the phoneme /p/ to P by the Roman alphabet, and the phoneme /v/ to B by the Cyrillic alphabet. Compare HAPEB to a nonword of the first type, such as BEMAP. BEMAP has all but one letter in common with a real word (BETAP), and all but one phoneme in common with this real word. HAPEB similarly has all but one letter in common with a real word (HAПEB). However, HAPEB has all phonemes in common with a real word, if both alphabets apply. If lexical access is visual, then false positives to BEMAP and HAPEB should be few and equal. In contrast, if lexical access is based on phonemic descriptions computed prelexically, and if the computation is analytic, assigning as many phonemes per letter as permitted, then false positives should be greater for HAPEB. Experimentation shows that BEMAP and its unambiguous nonword control VEMAR generate the same small number of false positives (about 2-3%). When preceded by a neutral context word, HAPEB generated 31% false positives; when preceded by an associate of /napev/, namely, MELODIJA ("melody"), HAPEB generated 55% false positives (Lukatela, Turvey et al., 1989). The latencies of the false positives in the neutral and associative contexts were 890 msec and 779 msec, respectively. Further, the false positive "yes" response times to HAPEB-type nonwords and correct "yes" response times to their corresponding control words (e.g., NAPEV) were closely similar, 824 msec to 795 msec. The same pattern of results with respect to BETAP and VEPAR, and HAPEB and NAPEV, held for naming (Lukatela, Feldman et al., 1989; Lukatela, Turvey, Feldman, Carello, & Katz, 1989). These results are understandable if (a) a phonologically ambiguous nonword can give rise automatically to all of the phonological representations that its letter structure permits, (b) for some phonologically ambiguous nonwords one of the automatically generated phonological representations corresponds to a word, and (c) the lexicon is accessed through phonological representations.

The conclusion to be drawn from the English and Serbo-Croatian data summarized in the preceding paragraphs is that phonological mediation plays a significant role in lexical access. It is the case, however, that the supportive data from English language studies remains limited; the majority of the evidence provided above was from Serbo-Croatian studies, and most of it was obtained with tasks exploiting the peculiarities of the Serbo-Croatian orthography. Our goal in the present article is to provide a further line of support within the English language for this core hypothesis of the dual process theory. The experiments reported constitute a return to the pseudohomophony manipulation introduced by Rubenstein et al. (1971); they exploit, therefore, a feature of the English orthography that has no counterpart in the phonetically precise Serbo-Croatian orthography. We place the pseudohomophony manipulation in the context of the associative priming phenomenon, and we use naming rather than lexical decision as the response measure. Associative priming is the improvement in word identification that follows from preceding the target word by an associate, for example, TABLE-CHAIR, where the control is provided by the same target word following a nonassociate, for example, NOVEL-CHAIR. This effect is said to arise from the connections within the internal lexicon. From the perspective of evaluating the nature of lexical access, naming has the advantage that, unlike lexical decision, it does not necessitate post-lexical processes. In signal detection terminology, whereas lexical decision involves sensitivity and bias, naming involves only sensitivity; it does not entail a decision that requires a criterion (Seidenberg, Waters, Sanders, & Langer, 1984).

Consider the task of naming the pseudohomophone CHARE following the "associate" TABLE. If TABLE's phonology is computed pre-lexically, and this representation is used to access the lexicon, then TABLE would activate fully its lexical representation /table/ and activate partially, through lexical interconnections, the associated representation /chair/. The preactivation of /chair/ could benefit the naming of CHARE in two ways. The first way is based on the reasonable assumption that the prelexically computed phonology is incomplete, for example, it would not specify stress. Consequently, to name a word, the full phonology must be retrieved from the lexicon. If CHARE is coded phonologically prior to lexical access, then it will contact /chair/ in the lexicon and, because of the associative priming of /chair/, retrieve the full phonology faster following TABLE than following a nonassociate. The second way CHARE could benefit from the priming of /chair/ is based on the assumption that the prelexical computation of phonology and the accessing of lexical phonology ordinarily interact in the determination of a letter string's name. If so, then CHARE will benefit from the prior presentation of TABLE through the feedback from the preactivated lexical item /chair/ to the process by which CHARE's phonology is assembled. (The notions of assembled phonology and lexical phonology are not mutually exclusive, and in interpretations of Serbo-Croatian word naming the need for both sources of phonological information has been repeatedly underscored [Carello, Lukatela, & Turvey, 1988; Lukatela & Turvey, 1987; Lukatela, Turvey et al., 1989; Lukatela & Turvey, 1990a]. An analogous, though weaker version, of the same idea has been advanced for English by McCann and Besner [1987].) Arguments of a similar kind apply to the even more interesting case in which the pseudohomophone is in the role of the prime, for example, TAYBLE-CHAIR. According to the phonological mediation hypothesis, the context stimulus TAYBLE will activate the lexical entry /table/ and, by lexical interconnections, the associate /chair/. As a result, naming CHAIR will benefit from the prior processing of TAYBLE.

In sum, the phonological mediation hypothesis of dual process theory makes the following predictions: (a) a target pseudohomophone will be named faster in the context of an associate of the pseudohomophone's source word than in an unrelated context, and (b) a target word will be named faster in the context of a pseudohomophone whose source word is an associate of the target word than in the context of a pseudohomophone whose source word is unrelated to the target word. Experiments 1 and 2 test prediction (a), Experiments 3 and 4 test prediction (b).

EXPERIMENT 1

The first experiment uses stimuli constructed in the manner of McCann and Besner (1987) in their demonstration that pseudohmophones are named faster than ordinary nonwords. The nonpseudohomophone controls (e.g., THARE) were distinguished from the pseudohomophones (e.g., CHARE) by a single initial letter. The experiment compared target naming latencies for TABLE-CHARE, NOVEL-CHARE, TABLE-THARE, and NOVEL-THARE pairs.

Method

Subjects. The participants in the experiment were 24 students from the Department of Psychology at the University of Connecticut. A subject was assigned to one of four groups to give a total of six subjects per group.

Materials. A set of 60 semantically related word pairs (e.g., TABLE-CHAIR) was generated on the basis of a high associativity rating in Palermo and Jenkins (1964), and in a tabulation of associative norms collected at the University of Connecticut. In each context-target pair the target word was replaced by its pseudohomophone to produce a set of 60 word-pseudohomophone pairs (TABLE-CHARE). With the same pseudohomophone targets, another set of 60 associatively unrelated pairs (NOVEL-CHARE) was also generated.

A third set of 60 word-nonword pairs (e.g., TABLE-THARE) was generated by replacing the first (or, in some cases, the second) grapheme of each pseudohomophone (CHARE) with another legal grapheme (THARE). A fourth set of 60 word-nonword pairs was assembled by combining the associatively unrelated context word (GULF) with the non-pseudohomophone target (THARE).The complete set of experimental materials is presented in Appendix 1. In addition, a foil set of 48 word-word pairs was selected from a different word list. In one half of the foil pairs the context and target word were associatively related and in the other half they were unrelated.

Four counterbalanced lists of stimuli pairs were prepared. In each group, each subject saw 15 instances of each stimulus pair. For example, in one group the subject would see 15 FOOT-HANNED (associated pseudohomophone derived from HAND) pairs, 15 HAMMER-LALE (associated nonpseudohomophone derived from NAIL) pairs, 15 FACT-KWEEN (nonassociated pseudohomophone derived from QUEEN) pairs, and 15 DRESS-SUDDER (nonassociated nonpseudohomophone derived from BUTTER) pairs. Additionally each subject saw 24 related word-word pairs (GOLF-BALL),and 24 unrelated word-word pairs (FOOL-BALL).

Design. The major constraint on the design was that a given subject never encountered a given target more than once. There were four (2 × 2) stimulus types (associative relation × target homophony). Each subject was presented with 15 experimental stimulus pairs from each of the four types and, in addition, with 48 foils for a total of 108 stimulus pairs. The experimental sequence was divided into quarters, with a brief rest in between. Stimulus types were ordered pseudorandomly within each quarter. Experimental sequence was preceded by a practice sequence of 32 stimulus pairs.

Procedure. Subjects, run one at a time, sat in front of the CRT of an Apple IIe computer in a dimly lit room. A fixation point was centered on the screen. On each trial, there was a brief auditory warning signal after which a letter string appeared for 500 ms above the fixation point. After a 100 ms interstimulus interval, another letter string appeared below the fixation point for 500 ms. The subject was told that he or she would be viewing word-word and word-nonword pairs, and that some of the nonwords, when pronounced, would sound like English words. Subjects were required to pronounce each target letter string as quickly and as distinctly as possible. In all conditions, latencies from the onset of the target to the onset of the response were measured by a voice operated trigger relay. Naming was considered to be erroneous when the pronunciation included a phoneme not specified by allowable grapheme-to-phoneme correspondences in English, the pronunciation was not smooth (i.e., the subject hesitated after beginning to name), or the response was not loud enough to trigger the voice key. To ensure that subjects were reading the contexts, the subject was asked occasionally (on less than 5% of total trials) by a computer message to report orally both words (only some of the foil word-word pairs were controlled) after the target word was named. If the naming latency was longer than 1000 ms, a message appeared on the screen requesting the subject to name more quickly. All latencies, including those longer than

1000 ms, were correctly stored in the computer memory.

Results and Discussion

For each subject, naming latencies more than two standard-deviation units above or below the subject's mean in all conditions were considered as errors. (This procedure was followed for all of the reported experiments.) The mean latencies of each group of subjects, standard deviations, and errors for related and unrelated contexts are presented in Table 1.

Table 1. *Mean naming latencies L (in ms) and error rate ER (in %), with the corresponding standard deviations by subjects and by items in Experiment 1.*

Target	Context-target Relation			
	Related		Unrelated	
	L	ER	L	ER
Pseudohomophone				
	600[a]	5.83	621	11.94
	81[b]	6.61	82	9.63
	68[c]	12.21	80	17.92
Non-Pseudohomophone				
	661	7.22	646	5.28
	94	5.17	109	7.35
	78	10.79	90	9.45

[a] mean
[b] standard deviation by subjects
[c] standard deviation by items

A 2 × 2 analysis of variance was conducted on the latency data with factors of associativeness, and homophony (pseudohomophones vs. non-pseudohomophones). Homophony was significant (pseudohomophones = 610 ms vs. non-pseudohomophones = 653 ms) by subjects $F(1, 23) = 35.53$, $p < .001$, and by stimuli, $F(1,59) = 34.04$, $p < .001$. Associativeness was not significant as a main effect (related = 630 ms vs. unrelated = 633 ms), but interacted with homophony (associative effect for pseudohomophones = +21 ms vs. associative effect for non-pseudohomophones = -15 ms) significantly by subjects, $F(1, 23) = 14.83$, $p < .001$, and by stimuli, $F(1,59) = 4.33$, $p < .05$.

Because of the associativeness with pseudohomophony interaction, separate analyses were performed on the pseudohomophone data, and the nonpseudohomophone data. Thus, associativeness for the pseudohomophone latencies (related = 600 ms vs. unrelated = 621 ms) proved significant by subjects, $F(1, 23) = 17.38$, $p < .001$, and by stimuli, $F(1, 59) = 3.88$, $p < .05$.

A 2 × 2 error analysis showed no significant main effects. There was, however, a significant interaction of associativeness and homophony (associative effect for pseudohomophone errors = + 6.11% vs. associative effect for non-pseudohomophone errors = - 1.94%) by subjects, $F(1, 23) = 8.23$, $p < .01$, and by stimuli, $F(1, 59) = 7.19$, $p < .01$. A separate analysis for pseudohomophone error data showed that associativeness (related = 5.83% vs. unrelated = 11.94%) was significant by subjects and stimuli, $F(1, 23) = 6.84$, $p < .01$, $F(1, 59) = 5.07$, $p < .05$, respectively. For the non-pseudohomophone data, the main effect of associativeness (related = 661 ms vs. unrelated = 646 ms) was insignificant in the latency data, $F(1, 23) = 3.74$, $p > .05$ by subjects, and $F(1, 59) = 1.42$, $p > .05$, by stimuli. Associativeness (related = 7.22% vs. unrelated = 5.28%) was also insignificant in the non-pseudohomophone error data, $F(1, 23) = 1.62$, $p > .05$ by subjects, and $F < 1$ by stimuli.

In sum, CHARE was named faster and with fewer errors (as defined above) when the context was TABLE than when the context was NOVEL. The naming of THARE, in contrast, showed statistically no differential sensitivity to TABLE and NOVEL.

EXPERIMENT 2

Although the outcome of the first experiment was straightforward, questions might be raised about (a) the high error rates for nonrelated pseudohomophone targets, and (b) the manner in which the control nonwords were created. With respect to (a) one may note that some of the pseudohomophones (for example, KLOWD, HAHT, SMUTHE, ... standing for CLOUD, HOT, SMOOTH, ... , respectively) in Experiment 1 (see Appendix 1) did not constrain the reader to the expected homophonic pronunciation and, evidently, left room for various phonological interpretations. With respect to (b) one may object that the graphemic structure of non-pseudohomophones (e.g., THARE) was less like the source word CHAIR than was the graphemic structure of pseudohomophones (e.g., CHARE). Consequently, the differences obtained in Experiment 1 could be attributed wholly or in part to the graphemic advantage of the pseudohomophones over their controls. The second experiment addressed this potential criticism. If CHARE was the pseudohomophone, then the new control would be CHARK, that is, a letter string that has the same letter overlap with the source word CHAIR as does the corresponding pseudohomophone. More precisely speaking, each "pseudohomograph" control and the corresponding pseudohomophone would share the same number of letters with the source word in terms of both exact letter positions and total number of letters. Because this manipulation required an almost new basic set of 64 source pairs (different from that in Experiment 1), the second experiment also tests the generality of the preceding results. Finally, in the second experiment one set of words served as both associated and nonassociated contexts in contrast to the first experiment in which different sets of words functioned in these roles. This additional manipulation guarded against uncontrolled effects due to differences in average familiarity of the context stimuli.

Method

Subjects. The participants in the experiment were 24 students from the Department of Psychology at the University of Connecticut. A subject was assigned to one of four groups to give a total of six subjects per group. The subjects did not participate in Experiment 1.

Materials. The stimuli (see Appendix 2) were generated from a set of 64 associatively related word pairs was generated on the basis of a high associativity rating in Palermo and Jenkins (1964), and in a tabulation of associative norms collected at the University of Connecticut. In addition, a list of different words, which can be conveniently transcribed as pseudohomophones, was compiled in part from the source words used by Perfetti, Bell, and Delaney (1988), and in part from the authors' own work. The list was presented to a pool of 60 undergraduate students of Psychology at the University of Connecticut. Each student in the pool was asked to read through the list on a word by word basis and after each word to write down on a prepared form the first three words that may have come to his/her mind. The appropriate associates were picked up to create new associatively related word-word pairs that were paired to make 32 "related word quadruples" (e.g., BITE-TEETH; CIDER-JUICE). Within each related quadruple the mutual substitution of context words produced a new "unrelated word quadruple" of two associatively unrelated pairs (e.g., CIDER-TEETH, BITE-JUICE). In each context-target pair the target word was replaced by its pseudohomophone to produce a set of 64 associatively related word-pseudohomophone pairs or 32 related pseudohomophone quadruples (BITE-TEATH; CIDER-JOOCE). With the same pseudohomophone targets, a second set of 32 associatively unrelated pseudohomophone quadruples (e.g., CIDER-TEATH; BITE-JOOCE) was also generated.

A third set of 64 "associated" word-pseudohomograph pairs or 32 related pseudohomograph quadruples (e.g., BITE-TERTH; CIDER-JORCE) was generated, whereby each pseudohomograph target shared the same letters as the pseudohomophone target, in position and out of position, with the source word (e.g., TEETH, JUICE, respectively). A fourth set of 32 unrelated pseudohomograph quadruples was assembled by mutually substituting the two context words in each quadruple (e.g., CIDER-TERTH; BITE-JORCE). With regard to all pseudohomophones and pseudohomographs used in this and in the next experiment a word of caution might be necessary: they met an arbitrary criterion of lexicality for a letter sequence. Namely, the pseudohomophones and pseudohomographs should have no entry in the Word Frequency Book by Carroll, Davies, and Richman (1971) which is based on the American Heritage Intermediate Corpus of approximately 5,000.000 words (tokens).

In order to prevent any special response strategy in rapid naming, a foil set of 32 associated word-word pairs (e.g., SOFT-HARD) and a corresponding (i. e., having the same target words) foil set of 32 unrelated word-word pairs (e.g., CHESS-HARD) were selected from a different word list. In addition, a third foil set of 32 word-nonword pairs was also created. The nonword targets were easily pronounceable, legal English letter strings.

As before, four counterbalanced lists of stimuli pairs were prepared. In each group, each subject saw eight quadruples with 16 instances of each stimulus pair. For example,in one group the subject would see 16 BITE-TEATH (associated pseudohomophone derived from TEETH) pairs, 16 DEAF-MERT (associated pseudohomograph derived from MUTE) pairs, 16 MOAN-FYVE (unrelated pseudohomophone derived from FIVE) pairs, and 16 WEAR-BUSHER (unrelated pseudohomograph derived from BUTTER) pairs. Additionally each subject saw 16 related word-word pairs (SOFT-HARD), 16 unrelated work-word pairs (OLIVE-STRICT), and 32 word-nonword pairs (ROSE-SHAIG). In total, an experimental list consisted of 128 stimulus pairs.

Design and procedure. These were the same as in Experiment 1, except for the fact that the context and target presentation time both were reduced to 400 ms each, leaving the interstimulus interval unchanged (100 ms).

Results and Discussion

The mean latencies, standard deviations, and errors for pseudohomophones and pseudohomographs in associatively related and unrelated contexts are presented in Table 2.

Table 2. *Mean naming latencies L (in ms) and error rate ER (in %), with the corresponding standard deviations by subjects and by items in Experiment 2.*

Target	Context-target Relation			
	Related		Unrelated	
	L	ER	L	ER
Pseudohomophone				
	557[a]	3.39	578	3.91
	70[b]	4.51	79	4.44
	59[c]	8.84	54	8.56
Pseudohomograph				
	582	2.86	574	3.91
	78	4.87	78	4.44
	48	7.08	50	10.80

[a]mean
[b]standard deviation by subjects
[c]standard deviation by items

A 2 × 2 analysis of variance was conducted on the latency data with factors of associativeness, and homophony (pseudohomophones vs. pseudohomographs). Homophony was significant (pseudohomophones = 567 ms vs. pseudohomographs = 578 ms) by subjects $F(1,23) = 7.31$, $p < .01$, but not by stimuli, $F(1,63) = 3.57$, $p > .05$. Associativeness was not significant as a main effect (related = 569 ms vs. unrelated = 575 ms), but it interacted with homophony (associative effect for pseudohomophones = +20 ms vs. associative effect for pseudohomographs = - 5 ms) significantly both by subjects, $F(1, 23) = 19.72$, $p < .001$, and stimuli, $F(1, 63) = 5.15$, $p < .05$.

Because of the associativeness with homophony interaction, separate analyses were performed on the pseudohomophone data, and the pseudohomograph data. Thus, associativeness for the pseudohomophone latencies (related = 557 ms vs. unrelated = 578 ms) proved significant by subjects, $F(1, 23) = 10.91$, $p < .01$, and by stimuli, $F(1, 63) = 5.26$, $p < 0.5$. A 2 × 2 error analysis showed no significant effects and no interaction.

Considering the pseudohomograph data, the main effect of associativeness was not significant for latencies (related = 582 ms vs. unrelated = 574 ms)—by subjects, $F(1, 23) = 2.06$, $p > .1$, by stimuli, $F(1, 63) = 0.2$, $p > .1$—nor was it significant for errors (all $F < 1$). The mean naming latency for associatively related pseudohomophones (557 ms) was reliably faster than for associatively related pseudohomographs (582 ms), both by subjects, $F(1, 23) = 18.17$, $p < .001$, and by stimuli, $F(1, 63) = 7.18$, $p < .01$.

The obtained results suggest an associative priming of pseudohomophones, and no contextual effect for the pseudohomographic controls. The magnitude of the obtained contextual effect (+21 ms) compared favorably with the magnitude obtained on the foil set of 32 word-word stimulus pairs (HARD-SOFT = 525ms vs. CHESS-SOFT = 540 ms).

Experiment 3

Although the results of Experiments 1 and 2 can be discussed in terms of the automatic priming of /chair/, and of the prelexical phonological processes by which CHARE's name is assembled, other interpretations are possible. In particular, given the challenge of naming nonwords rapidly, a subject might have adopted a conscious strategy of generating on each trial a word associated with the context in order to help in the pronunciation of the nonword target. The fact that 25% of the pairs were of the GOLF-BALL type and 16% were of the TABLE-CHARE type might have encouraged such a strategy.

The third experiment examined the efficacy of a pseudohomophone as an associated prime. Specifically, the experiment compared naming latencies to CHAIR when presented subsequent to TAYBLE and TARBLE in an experimental setting in which TARBLE-CHAIR pairs accounted for only about 8% of all pairs. This experiment, therefore, militates against the conscious association-pronunciation strategy identified above, and should provide thereby a purer test of the phonological mediation hypothesis than the preceding two experiments. According to the phonological mediation hypothesis, TAYBLE will activate the lexical entry /table/, resulting in the priming of that entry's associate, /chair/. The context TARBLE, although visually similar to TAYBLE, will not activate the lexical item /table/ and, therefore, not prime /chair/. Consequently, naming CHAIR should be faster in the context TAYBLE than in the context TARBLE.

The design of the third experiment is significant in an additional way. If TAYBLE-CHAIR is faster than TARBLE-CHAIR, then it would emphasize the prelexical/lexical nature of pseudoassociative priming. An effect arising from the combination of naming as the dependent measure and pseudohomophones as primes would pose difficulties for any post-lexical account. The consensus is that naming involves few, if any, post-lexical checks, and that associative priming is purely lexical (Balota & Chumbley, 1985; Carello et al., 1988; Seidenberg et al., 1984)

Method

Subjects. The participants in the experiment were 24 students from the Department of Psychology at the University of Connecticut. A subject was assigned to one of four groups to give a total of six subjects per group. None of the subjects had participated in the previous experiments.

Materials. The stimuli (see Appendix 3) were generated from the same general list which was used in Experiment 2, though not all of the word-word pairs in this experiment were necessarily identical to those in Experiment 2. A basic set of 64 associatively related word pairs that were paired to make 32 "related word quadruples" (e.g., BITE-TEETH; HATE-LOVE). Within each related quadruple the mutual substitution of context words produced a new "unrelated word quadruple" of two associatively unrelated pairs (e.g., HATE-TEETH, BITE-LOVE). In each context-target pair the context word was replaced by its pseudohomophone to produce a set of 64 associatively related pseudohomophone-word pairs or 32 related pseudohomophone quadruples (BIGHT-TEETH; HAIT-LOVE). With the same word targets, a second set of stimuli which comprised 32 associatively unrelated pseudohomophone quadruples (e.g., HAIT-TEETH; BIGHT-LOVE) was also generated.

A third set of 64 "associated" pseudohomograph-word pairs which comprised 32 related pseudohomograph quadruples (e.g., BIRET-TEETH; HANT-LOVE) was generated, whereby each pseudohomograph context shared the same letters as the pseudohomophone context, in position and out of position, with its source word (e.g., BITE, HATE, respectively). A fourth set of 32 "unrelated" pseudohomograph quadruples was assembled by mutually substituting the two context words in each quadruple (e.g., HANT-TEETH; BIRET-LOVE). Summed bigram frequency (Mayzner & Tresselt, 1965) was computed for each pseudohomophone and pseudohomograph; in the majority of pseudohomographs (i.e., 49 out of 64) the summed bigram frequency was higher than in the corresponding pseudohomophones. In addition, a comparison set of 32 associated word-word pairs (e.g., BOY-GIRL) and another set having the same target words but unrelated context words (e.g., GULF-GIRL) were selected from a different word list. The targets words in the comparison set were selected to have approximately the same average frequency of occurrence as the target words in the four experimental sets. In order to prevent any special response strategy in rapid naming, a foil set of 64 nonword-nonword pairs was assembled. In the same vein, to discourage further a subject's guessing strategy, a foil set of 32 pseudohomophone-nonword pairs was also created. All nonwords were easily pronounceable, legal English letter strings.

As in the preceding experiments, four counterbalanced lists of stimuli pairs were prepared. In each group, each subject saw eight quadruples with 16 instances of each stimulus pair. For example, in one group the subject would see 16 BIGHT-TEETH (associated pseudohomophone context derived from BITE) pairs, 16 DELF-MUTE (associated pseudohomograph context derived from DEAF) pairs, 16 MONE-FIVE (unrelated pseudohomophone context derived from MOAN) pairs, and 16 WAUR-BUTTER (unrelated pseudohomograph context derived from WEAR) pairs. Additionally each subject saw 16 related word-word pairs (BOY-GIRL), 16 unrelated work-word pairs (SKULL-LAW), 32 pseudohomophone-nonword pairs (LOGIK-SHAIG), and 64 nonword-nonword pairs (TINFER-PROLIT). The foil sets were used to counter the development of biases, such as (a) always looking for an associative relation, (b) making predictions about targets based on the sound of the context, and (c) trying to make contexts sound like words. In total, an experimental list consisted of 192 stimulus pairs.

Design. The design was similar to that in Experiments 1 and 2, except for an increase in the number of foil stimuli pairs. This increase produced a low ratio (1 : 3) in the presentation of associatively related relative to associatively unrelated stimulus pairs within each quarter. The experimental sequence was preceded by a practice sequence of 32 stimulus pairs.

Procedure. The procedure was the same as that in Experiment 2.

Results and Discussion

The mean latencies, standard deviations, and errors for pseudohomophones and pseudohomographs in associatively related and unrelated contexts are presented in Table 3. A 2 × 2 analysis of variance was conducted on the latency data with factors of associativeness, and the homophony of the context (pseudohomophone context vs. pseudohomograph context). Associativeness as a main effect (related = 530 ms vs. unrelated = 538 ms) was significant only by subjects, $F(1, 23) = 4.33$, $p < .05$, but not by stimuli, $F(1. 63) = 1.66$, $p > .1$. Main effect of context homophony was not significant, $F(1, 23) = 1.43$, $p > .1$, $F(1, 63) = 0.96$, $p > .1$, by subjects and stimuli, respectively. However, context homophony interacted with associativeness (associative effect for pseudohomophone context = +20 ms vs. associative effect for pseudohomograph context = - 4 ms). The interaction was significant both by subjects, $F(1, 23) = 17.84$, $p < .001$, and by stimuli, $F(1, 63) = 4.35$, $p < .05$.

Table 3. *Mean naming latencies L (in ms) and error rate ER (in %), with the corresponding standard deviations by subjects and by items in Experiment 3.*

Context	Context-target Relation			
	Related		Unrelated	
	L	ER	L	ER
Pseudohomophone context				
	522[a]	1.82	542	0.78
	45[b]	2.90	47	2.11
	41[c]	5.24	46	3.73
Pseudohomograph context				
	538	3.13	534	1.04
	53	4.12	49	2.38
	51	7.63	41	3.60

[a] mean
[b] standard deviation by subjects
[c] standard deviation by items

The associativeness by context homophony interaction motivated separate analyses on the pseudohomophone context pairs and the pseudohomograph context pairs. Thus, the effect on naming latency of associativeness when contexts were pseudohomophones (related = 522 ms vs. unrelated = 542 ms) proved significant by subjects, $F(1, 23) = 19.07$, $p < .001$, and by stimuli, $F(1, 63) = 6.76$, $p < 0.01$.

A 2 × 2 error analysis displayed a significant main effect of associativeness (related = 2.47% vs. unrelated = 0.91%), $F(1, 23) = 5.75$, $p < .05$, $F(1, 63) = 7.28$, $p < .01$. The main effect of context homophony was not significant, $F(1, 23) = 1, 68$, $p > .1$, $F(1, 63) = 1.54$, $p > .1$; neither was the interaction between context homophony and associativeness, $F < 1$. A similar analysis on the context pseudohomograph data revealed that the main effect of associativeness was not significant for latencies (related = 538 ms vs. unrelated = 534 ms), either by subjects or by stimuli ($F < 1$ in both cases).

A further observation was that the mean naming latency for word targets preceded by associatively related pseudohomophones (557 ms) was reliably faster than the mean naming latency for word targets preceded by associatively related pseudohomographs (582 ms), both by subjects, $F(1, 23) = 18.17$, $p < .001$, and by stimuli, $F(1, 63) = 7.18$, $p < .01$.

How efficient are pseudohomophones as associative primes compared to real word contexts? An approximate evaluation is provided by an ANOVA (with factors of associativeness and context lexicality) conducted on the target latencies of the experimental set of 32 pseudohomophone-word pairs (e.g., TAYBLE-CHAIR) and the set of 32 word-word pairs (BOY-GIRL). The analysis was within subjects and between items. The evaluation is only approximate because of the different target items and different associative strengths in the comparison pairs. Only the main effect of associativeness (related = 521 ms vs. unrelated = 540 ms) was significant, $F(1, 23) = 38.87$, $p < .001$, $F(1, 62) = 10.24$, $p < .01$, by subjects and by stimuli, respectively. The important effects of context lexicality (pseudohomophones = 532 ms vs. words = 529 ms), and of associativeness by context lexicality (associative effect by pseudohomophone contexts = + 20 ms vs. associative effect by word contexts = + 19 ms) were not significant, $F < 1$ in both cases.

Finally, we note that given the design of the experiment, a strategy of looking for associations was not encouraged (only 1/6 of all stimulus pairs were associated words), neither was the assembling of a name for the prime (only targets had to be named). Despite these discouraging features of the experimental design, TAYBLE primed CHAIR suggesting that neither the effect of association nor the effect of phonology were optional.

EXPERIMENT 4

In Experiment 3 the stimulus onset asynchrony (SOA) between context and target was 500 ms. With respect to the trials with nonword contexts, the delay from context to target may have allowed subjects the opportunity to construct the context's phonology and then to determine consciously whether or not there was a word with which it was homophonic. That is to say, the source of TAYBLE's advantage over TARBLE in the priming of CHAIR may have resided in strategic processes of a post-lexical nature rather than automatic processes of a pre-lexical and lexical nature. The inclusion in Experiment 3 of a relatively small proportion of TAYBLE-CHAIR pairs (8.3% of the total number of pairs) was for the purpose of minimizing conscious guessing strategies of the preceding kind, but the possibility remains that this protective measure was not sufficient. In the fourth experiment the conditions of priming by a nonword homophonic with an associate of the target were repeated with SOAs of 280 ms and 500 ms. A much reduced delay of target relative to the context ought to restrain the use of conscious strategies, or at least render them less effective. Consequently, if the priming effect of TAYBLE is strategic, then there should be an interaction between onset asynchrony and the associative relation between context and target; that is, relative to TARBLE, TAYBLE should prime CHAIR less with SOA = 280 ms than with SOA = 500 ms.

In addition to examining the effect of SOA, the fourth experiment was designed to provide an assessment of the degree to which priming a target word by a nonword homophonic with an associate of the target is comparable to priming a target word by an associated word. In the third experiment the relative efficacy of pseudohomophonic priming had been evaluated statistically by comparing pairs such as TAYBLE-CHAIR with pairs such as BOY-GIRL. Neither the lexical status of the context nor the associativeness by context lexicality interaction were significant. In the fourth experiment, the comparison is made across pairs that have the same target and the same associative strength, for example, TAYBLE-CHAIR vs. TABLE-CHAIR. Evidence that pseudohomophonic nonwords prime as well as words will be looked for in the latency differences between (a) CHAIR preceded by TAYBLE and BLOO and (b) CHAIR preceded by TABLE and BLUE. If pseudohomophones prime equally as well, then the advantage of TAYBLE over BLOO in (a) should be of the same magnitude as the advantage of TABLE over BLUE in (b), and if the pseudohomophonic priming effect witnessed in the preceding experiments is due to non-strategic processes, then this equation should hold equally over both the short (280 ms) and long (500 ms) SOAs.

Method

Subjects. The participants in the experiment were 60 students at the University of Connecticut. Subjects received either credit for participating as part of the course requirements in Introductory Psychology or were paid $5. A subject was assigned to one of two main groups (distinguished by SOA), and within a group to one of six subgroups to give a total of five subjects per subgroup. None of the subjects had participated in the previous experiments.

Materials. There were six word sets (see Appendix 3). The first set (also the base set) consisted of 96 associatively related word pairs that were paired to make 48 "related word quadruples" (e.g., BITE-TEETH; HATE-LOVE). From this first set, five other sets of 96 pairs were generated in the manner described in Experiment 3: Set 2 (unrelated word) represented by BITE-LOVE, Set 3 (related pseudohomophone) represented by HAIT-LOVE, Set 4 (unrelated pseudohomophone) represented by BIGHT-LOVE, Set 5 (related pseudohomograph) represented by HANT-LOVE, and Set 6 (unrelated pseudohomograph) represented by BIRET-LOVE.

Six counterbalanced lists of stimuli pairs were prepared. In each subgroup, each subject saw 16 pairs from of each stimulus set (for a total of 96), plus a foil set of 64 nonword-nonword pairs, a foil set of 32 pseudohomophone-nonword pairs (all nonwords were easily pronounceable, legal English letter strings), and a small set of eight nonrelated word-word pairs used for probing awareness of the context. The foil sets were used, as in Experiment 3, to counter the development of biases, such as (a) always looking for an associative relation, (b) making predictions about targets based on the sound of the context, and (c) trying to make contexts sound like words. In total, an experimental list consisted of 200 stimulus pairs.

Procedure. The stimuli were presented in a different manner from the previous experiments. Contexts and targets were fixed at 200 ms nominal duration. The interstimulus interval was either 300 ms or 80 ms to produce nominal SOAs of 500 ms and 280 ms, respectively. In further contrast to the preceding experiments, the context and target were presented at the same place (at the fixation point). The latter change in the presentation of stimuli was demanded by the fact that for SOA = 280 ms the time to shift fixation within the procedure used in Experiments 1-3 would be detrimental to performing the main task (e.g., the subject would learn to fixate on the area below the fixation point where the target appeared consistently). A major consequence of this change in manner of displaying the paired stimuli was the possibility for forward and/or backward masking, especially at SOA = 280 ms. In this regard, it is important to note that pilot work revealed that, for the nominal stimulus exposure times of 200 ms, nominal SOAs briefer than 280 ms (e.g., 230 ms) produced an unacceptable level of error in reporting the context of the probed pairs.

Results and Discussion

The mean latencies, mean errors, and standard deviations for words, pseudohomophones, and pseudohomographs in associatively related and unrelated contexts at both SOAs are presented in Table 4. A 2 × 3 × 2 ANOVA was conducted with factors of associativeness (within Ss), context (within Ss), and SOA (between Ss). The main effect of associativeness (associated = 547 ms vs. nonassociated = 557 ms) was significant, $F(1, 58) = 33.89, p < .001$ (by subjects), $F(1, 190) = 13.66, p < .001$ (by stimuli). Also significant was the main effect of context (word = 543 ms, pseudohomophone = 554 ms, pseudohomograph = 559 ms), $F(2, 116) = 25.37, p < .001$ (by subjects), $F(2, 380) = 16.33, p < .001$ (by stimuli). The main effect of SOA was not significant (long SOA = 542 ms vs. short SOA = 562 ms) by the subjects analysis, $F(1, 58) = 2.01, p > .05$ but it was significant by the stimuli analysis, $F(1, 190) = 36.63, p < .001$. Of all possible interactions, only the associativeness by context interaction (related-unrelated difference for words = 16 ms vs related-unrelated difference for pseudohomophones = 17 ms vs related-unrelated difference for pseudohomographs = -3 ms) was significant: $F(2, 116) =17.84, p < .001$ (by subjects), $F(2, 380) = 5.09, p < .007$ (by stimuli).

Table 4. *Mean naming latencies L (in ms) and error rate ER (in %), with the corresponding standard deviations by subjects and by items in Experiment 4.*

	Context-target Relation			
Context	Related		Unrelated	
	L	ER	L	ER
SOA = 500 ms				
Word				
	527[a]	2.29	545	1.25
	53[b]	3.84	54	2.54
	47[c]	7.61	45	4.87
Pseudohomophone				
	537	0.21	552	1.04
	47	1.14	47	2.37
	44	2.04	43	4.47
Pseudohomograph				
	549	1.25	544	0.83
	50	3.03	54	2.16
	46	5.67	43	4.02
SOA = 280 ms				
Word				
	543	0.83	556	1.46
	60	2.16	61	3.55
	50	4.02	48	5.23
Pseudohomophone				
	555	1.46	573	1.67
	54	3.91	59	3.26
	43	5.23	50	6.27
Pseudohomograph				
	572	0.42	571	0.83
	57	1.59	59	2.16
	49	2.87	47	4.02

[a] mean
[b] standard deviation by subjects
[c] standard deviation by items

Turning to the partial interactions, both that between words and pseudohomographs and that between pseudohomophones and pseudohomographs were significant: $F(1, 58) = 22.02, p < .001$(by subjects), $F(1, 190) = 7.70, p < .006$ (by stimuli) and $F(1, 58) = 28.14, p < .001$ (by subjects), $F(1,190) = 7.77, p < .006$ (by stimuli), respectively. The partial interaction between words and pseudohomophones was not significant—$F < 1$ in both the subjects and stimuli analyses. Importantly, the partial interaction between words and pseudohomophones was not significant at either SOA, all *F*s < 1, suggesting that the associativeness by context interaction was due solely to the behavior of the TARBLE-CHAIR/BLOM-CHAIR pairs relative to the word and pseudohomophone pairs.

This fourth experiment was designed to determine whether or not associative priming of word targets through pseudohomophones was dependent on the magnitude of time elapsing between context and target. Experiment 3 had used an SOA of 500 ms. The present experiment used an SOA of the same magnitude and one that was considerably shorter, viz., 280 ms. The analyses show that the efficacy of pseudohomophone associative priming was (a) indifferent to the SOA difference, and (b) of the same magnitude as word associative priming. For SOA = 500 ms, TABLE-CHAIR differed from its unrelated control by 18 ms (this planned comparison was significant by both subjects and stimuli, $F(1, 29) = 14.51, p < .001$ and $F(1, 95) = 8.46, p < .005$, respectively), TAYBLE-CHAIR differed from its unrelated control by 15 ms (significant by both subjects and stimuli, $F(1, 29)$

= 27.26, $p < .001$ and $F(1, 95) = 5.82$, $p < .02$, respectively), and TARBLE-CHAIR differed from its unrelated control by -5 ms (insignificant by both subjects and stimuli, $F(1, 29) = 3.12$, $p > .05$ and $F < 1$, respectively). For SOA = 280 ms, TABLE-CHAIR, TAYBLE-CHAIR, and TARBLE-CHAIR differed from their unrelated controls by 13 ms ($F(1, 29) = 24.29$, $p < .001$ and $F(1, 95) = 4.22$, $p < .05$), by 18 ms ($F(1, 29) = 17.98$, $p < .001$ and $F(1, 95) = 5.23$, $p < .03$), and by -1 ms ($F < 1$ for both subjects and stimuli), respectively. The implication is that one and the same mechanism underlies both pseudohomophone and word associative priming, and that pseudohomophone associative priming, therefore, is not due to a special (post-lexical) strategy.

There was a main effect of context lexicality in the sense that naming CHAIR was faster to the same degree following each of the two word contexts (TABLE and BLUE) than following each of the two pseudohomophone contexts (TAYBLE and BLOO). Clearly, this faster naming after word contexts has nothing to do with the mechanism of associative priming. It might be due to a post-lexical verification process in which the subject compares the assembled name with the initial representation of the stimulus prior to executing the response (see Paap, Newsome, McDonald, & Schvaneveldt, 1982). Suppose that the verification (match-mismatch) process is modulated over the short term by the kind of letter string just processed. Then the possibility can be entertained that processing a pseudohomophone raises temporarily the criterion for verification; as a consequence, the time to name an after-coming word is slowed. That pseudohomophone contexts may have exerted an inhibitory effect on word targets is suggested by the slower response to CHAIR following BLOO than following BLOM in the 500 ms SOA condition. The absence of a similar inhibitory effect in the 280 ms SOA condition would suggest that the raising of the verification criterion is not instantaneous.

GENERAL DISCUSSION

A large body of evidence favoring the idea that phonologic coding is the key process in printed word identification is available for Serbo-Croatian (see above and Lukatela & Turvey [1990a] for a review). The results of the experiments reported in the present article, especially Experiments 3 and 4, might be interpreted as adding to a growing body of evidence that phonological coding may be similarly dominant in printed word identification in English (Perfetti et al., 1988; Van Orden, 1987; Van Orden et al., 1988). In Experiments 1 and 2 we showed that a word prime can facilitate the naming of a nonword target if the nonword had the same phonology as an associate of the prime. In Experiments 3 and 4 we showed that a nonword prime can facilitate naming of a word target if its phonology is the same as that of an associate of the target. The general consensus is that associative priming—whether the mechanism be spreading activation, expectancy set, or resonant matching (Stone & Van Orden, 1989)—is a lexical process. In order for the observed associative effects to have occurred, there must have been lexical access. And given the homophony conditions under which these effects occurred, it must have been the case that the lexical representations accessed were phonological, and that the access was through phonology. It seems unlikely that an account of the data can be given in terms of a model that does not grant a central role to phonology in lexical access. Consider analogy models (e.g., Glushko, 1979), for example. TAYBLE would be assigned a phonology by analogy with its orthographic neighbors. To enhance the naming of CHAIR, this assembled phonological representation /table/ would have to affect the lexical representation of CHAIR. The minimal implication is that /table/ must function as a lexical access code, triggering lexical processes beneficial to the eventual naming of CHAIR.

In the light of the evidence presented here and elsewhere for phonological mediation, we are tempted to ask whether or not there is experimental support for another process separate from phonological mediation. A basic prediction of dual process theory is that irregular/exception words can only be processed by the direct access route because they cannot be coded by grapheme-phoneme correspondence rules. Van Orden et al. (1990) have argued that the logic by which irregular words constitute a priori evidence for direct processing depends on the validity of the claim that explicit, discrete rules map spellings to sounds. If the mapping is achieved in an alternative fashion, for example, by a continuous version of regularity (captured by the degree of covariation between orthographic and phonologic features across neighborhoods of words that are spelled similarly), then the prediction is not valid. These same authors also question critically the arguments for an independent direct process based on the phenomena of the word-superiority effect (McCusker et al., 1981) and successful reading by the deaf (McCusker et al., 1981; Seidenberg, 1985). As they note, there is considerable experimental evidence contrary to both arguments; phonological variables affect both the word-superiority effect (e.g., Chastain, 1981, 1984) and word identification by deaf readers (e.g., Hanson & Fowler, 1987).

One experimental result offered in favor of a separate direct access is the demonstration that, in a masking paradigm, phonologic priming fails to benefit target identification over and above orthographic priming (Evett & Humphreys, 1981). Thus, TILE primes FILE (phonologically similar/graphemically similar) no more than TOUCH primes COUCH (phonologically dissimilar/graphemically similar). The logic here is that if one can demonstrate orthographic priming without phonological priming, then one has evidence for an independent direct visual access. Van Orden et al. (1990) have suggested that the Evett and Humphrey's application of this logic is flawed in that the difference between the critical stimulus pairs was too small to produce a noticeable difference in identification. Moreover, there are other sources of data that suggest that TOUCH can affect the processing of COUCH detrimentally because of the phonological inconsistency between the items (Meyer, Schvanaveldt, & Ruddy, 1974). Although Hillinger (1980) provided evidence for strictly phonological priming in the lexical decision task, the observation has failed to replicate in further experiments with lexical decision (Martin & Jensen, 1988) and in experiments using naming latency (Peter, Turvey, & Lukatela, 1990). These same experiments, however, also failed to demonstrate any benefits to processing of a graphemically similar preceding stimulus. This confusing state of affairs with respect to forward priming experiments with English language materials contrasts sharply with the results from (a) similar forward priming experiments in Serbo-Croatian, and (b) backward priming experiments with English and Serbo-Croatian, which are in contradiction of an independent visual direct route.

With respect to (a), a large number of experiments demonstrate unequivocally that there is *no orthographic priming in lexical decision or naming over and above phonological priming* (Lukatela & Turvey, 1990a; Lukatela, Carello, & Turvey, 1990). In these experiments, the primes and targets consist of phonologically unique letter strings written either in different alphabets (e.g., prime is in Cyrillic, target is in Roman) and different cases, or in the same alphabet (both are in Cyrillic) and the same case. When they are in the same case the graphemically dissimilar/phonologically similar pairs never share more than one letter in common, and often they share none; the graphemically similar/phonologically similar pairs share all but one letter in common. In different cases and alphabets, there are no visually identical letter forms. Especially significant aspects of the results are that phonological similarity effects in both lexical decision and naming are independent of graphemic similarity; that phonological similarity expedites the naming of words and nonwords, and to the same degree; and that phonological similarity effect occurs even when the context is a masked nonword. The latter result in particular has special bearing on the issue of an independent direct visual route.

By hypothesis, once a word unit has been activated directly on the basis of its orthographic properties, phonological information about the word is made available. Suppose that the direct access is mandatory (automatic) and that the phonological mediation way is optional (nonautomatic), as has been frequently proposed. Then under masking conditions that minimize optional strategies, only the lexical way should work. On the basis of the preceding argument Humphreys, Evett, and Taylor (1982, p. 581) hypothesized that: "If phonological information is activated via a nonlexical route, a pseudohomophone priming effect should occur." Finding that the pseudohomophone condition (tial-TILE) in their masking experiment did not differ from its graphemic control condition (tirl-TILE), Humphreys et al. (1982) concluded that the lexical (direct visual access) way of deriving phonology is the only one of the two that is automatic and is probably the only way to activate phonological information. To the contrary, the successful demonstration of "pseudohomophone priming" under masking in Serbo-Croatian confirms the Humphreys et al. (1982) hypothesis and suggests, therefore, that it is the phonological access route that is automatic and (continuing their logic) is the only one available.

Let us now consider (b), the backward priming experiments. These entail the identification of briefly exposed target words under backward masking conditions with the following key features: The masks are nonwords, phonologically related or unrelated to the targets, and themselves followed by patterned stimuli to reduce their identification and, thereby, guessing strategies about target/mask relationships. The figural structure of the masks, and comparable intensities of the targets and masks, confine the masking effects on the targets to primarily central processes (Michaels & Turvey, 1979; Turvey, 1973).

Perfetti et al. (1988) argued that if phonology is computed automatically, then phonological similarity between the mask and target will reduce the interruption of central processing normally induced by the mask. They reasoned that a phonologically similar mask will reinforce the phonological information activated partially by the target. In contrast, a phonologically dissimilar mask will activate partially other phonological information. If it is the case that lexical activation follows from phonological information, then a target preceding a phonologically similar mask will be identified better than a target preceding a phonologically dissimilar mask. The idea is that lexical entries partially activated by a target will be activated further by a subsequent mask with common phonological properties. The outcomes of experiments by Naish (1980) and by Perfetti et al. (1988), both using native speakers/readers of English and English language materials, were in agreement with this prediction, as were the experiments of Lukatela and Turvey (1990b) using native speakers/readers of Serbo-Croatian and Serbo-Croatian materials. All of these experiments showed significantly higher levels of target identification for homophonous masking than for nonhomophonous masking. Put differently, none showed graphemic backward priming effects over and above phonological priming effects. With respect to Evett and Humphrey's (1981) logic, this outcome is contrary to that expected from an independent direct visual access.

Van Orden et al. (1990) have remarked that traditionally the best evidence cited in favor of the independent direct access hypothesis comes from experiments that did not yield predicted effects of stimulus phonology—a strategy of arguing form other-than-positive effects. With respect to the semantic category task, the direct access hypothesis predicts that the effects of homophony should be reduced as a function of the familiarity of the homophones (direct access is the exclusive route for familiar words). Contrary to this prediction, the effect of homophone familiarity was null (Van Orden, 1987; Van Orden et al., 1988). In the spirit of the other-than-positive effects strategy, this outcome should be interpreted in favor of phonological mediation and against direct visual access (Van Orden, in press; Van Orden et al., 1990).

In sum, despite the ubiquity of the claim for direct visual access the evidence is not strong, raising the possibility that the overarching hypothesis of Coltheart's (1978) original theory of two independent processes could well be false. In contrast, there is substantial and accumulating evidence, including that reported in the present article, for phonological mediation. We conclude by noting that, although discussions of differences among lexical access codes have usually been expressed in terms of activation times (e.g., orthographic features are earlier sources of constraint on lexical coding than phonologic features, or vice versa), other ways of expressing the differences are possible. One interesting, contemporary idea from dynamical interpretations of lexical access is that all codes are activated in parallel, but the phonological code is initially more coherent (less noisy, closer to its "attractor") than the others and acts as a dynamic frame that constrains other linguistic encodings that are initially less coherent (Van Orden et al., 1990). It remains to be seen whether or not this new version of the phonological mediation hypothesis, in conjunction with the new sources of data, can deflect the criticisms levied at earlier versions (e.g., Humphreys & Evett, 1985).

REFERENCES

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effect in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24*, 89-106.

Carello, C., Lukatela, G., & Turvey, M. T. (1988). Rapid naming is affected by association but not by syntax. *Memory & Cognition, 16*, 187-195.

Carrol, J. B., Davies, P., & Richman, B. (1971). *Word frequency book.* Boston: Houghton Mifflin.

Chastain, G. (1981). Phonological and orthographic factors in the word-superiority effect. *Memory & Cognition, 9*, 389-397.

Chastain, G. (1984). Phonological access to the mental lexicon in a target discrimination task. *Journal of General Psychology, 110*, 53-60.

Coltheart, M. (1978). Lexical access in simple reading task. In G. Underwood (Ed.), *Strategies of information processing*. London: Academic Press.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). New York: Academic Press.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. T. (1978). Phonological recoding and lexical access. *Memory & Cognition, 6*, 391-402.

Evett, L. J., & Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. *Quarterly Journal of Experimental Psychology, 33*, 325-350.

Feldman, L. B., Kostic, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "Basic orthographic syllabic structure" in a phonologically shallow orthography. *Psychological Research, 45*, 55 - 72.

Feldman, L. B., & Turvey, M.T. (1983). Word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 288-298.

Forster, K. I. & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627-635.

Fredriksen, J. R. & Kroll, J. F. (1976). Spelling and sound: approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 361-379.

Glushko, R. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 674-691.

Green, D. W. & Shallice, T. (1976). Direct visual access in reading for meaning. *Memory and Cognition, 4,* 753-758.

Henderson, L. (1982). *Orthography and word recognition in reading.* New York: Academic Press.

Hillinger, M. L. (1980). Priming effects with phonemically similar words: the encoding-bias hypothesis reconsidered. *Memory & Cognition, 8,* 115-123.

Humphreys, G. F., Evett, L. J. & Taylor, D. E. (1982). Automatic phonological priming in visual word recognition. *Memory & Cognition, 10,* 576-590.

Humphreys, G. W., & Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. *The Behavioral and Brain Sciences, 8,* 689-740.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition, 8,* 124 - 132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. (1978). Bi- alphabetical lexical decision. *Language and Speech, 21,* 142-163.

Lukatela, G., Carello, C., & Turvey, M. T. (1987). Lexical representation of regular and irregular inflected nouns. *Language and Cognitive Processes, 2,* 1-17.

Lukatela, G., Carello, C., & Turvey, M. T. (1990). Phonemic priming by words and pseudowords. *European Journal of Cognitive Psychology, 2,* 375-394.

Lukatela, G., & Turvey, M. T. (1987). Loci of phonological effects in recognizing words written in a shallow orthography. *Psychological Research, 49,* 139-146.

Lukatela, G., & Turvey, M. T. (1990a). Automatic and prelexical computation of phonology in visual word identification. *European Journal of Cognitive Psychology, 2,* 325-344.

Lukatela, G., & Turvey, M. T. (1990b). Phonemic similarity effects and prelexical phonology. *Memory & Cognition., 18,* 128-152.

Lukatela, G., Feldman, L.B., Turvey, M. T., Carello, C., & Katz, L (1989). Context effects in bi-alphabetical word perception. *Journal of Memory and Language, 28,* 214-236.

Lukatela, G., Turvey, M. T., Feldman, L. B., Carello, C., & Katz, L. (1989). Alphabet priming in bi-alphabetical word perception. *Journal of Memory and Language, 28,* 237-254.

Martin, R. C. & Jensen, C. R. (1988). Phonological priming in the lexical decision task: A failure to replicate. *Memory & Cognition, 16,* 505-521.

Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single letter and digram[**] frequency counts for various word length and letter position combinations. *Psychonomic Monograph Supplements, 1,* 13-32.

McCann, R. S., & Besner, D. (1987). Reading Pseudohomophones: Implication for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance,* 13, 14-24.

McCusker, L. X., Hillinger, M. L., & Bias, R. G. (1981). Phonological recoding and reading. *Psychological Bulletin, 89,* 217-245.

Meyer, D. E., Schvaneveldt, R. W. & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word recognition. *Memory & Cognition,* 2, 309-321.

Michaels, C. F., & Turvey, M. T. (1979). Central sources of visual masking: Indexing the structures supporting seeing at a single brief glance. *Psychological Research, 41,* 1-61.

Naish, P. (1980). The effects of graphemic and phonemic similarity between targets and masks in a backward visual masking paradigm. *Quarterly journal of Experimental Psychology, 32,* 57-68.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model fro letter and word recognition: The word superiority effect. *Psychological Review, 89,* 573-594.

Perfetti, C. A., Bell, L. C., & Delaney, S. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59-70.

Peter, M., Turvey, M. T., & Lukatela, G. (1990). Phonological priming: Failure to replicate in the rapid naming task. *Bulletin of the Psychonomic Society, 28,* 389-392.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). An evidence for phonemic reading in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 10,* 645-654.

Seidenberg, M.S., Waters, G. S., Sanders, M. A., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition, 12,* 315-388.

Stone, G. O., & Van Orden, G. C. (1989). Are words represented by nodes? *Memory & Cognition, 17,* 511-524.

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from masking with patterned stimuli. *Psychological Review, 80,* 1-52.

Turvey, M. T., Feldman, L., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and reading* (pp.81-89). Hillsdale, NJ: Erlbaum.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory & Cognition, 15,* 181-198.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 371-385.

Van Orden, G. C. (in press). Phonologic mediation is fundamental to reading. In D. Besner & G. Humphreys (Eds.), Basic processes in reading: *Visual word recognition. Hillsdale,* NJ: Erlbaum.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488-522.

FOOTNOTES

[*] *Journal of Experimental Psychology: Human Perception and Performance,* in press.

[†] University of Belgrade, Belgrade, Yugoslavia.

[††] Also University of Connecticut, Storrs.

APPENDIX 1

Target source words (SW), pseudohomophones (PH), non-pseudohomophones (Non-PH), and the corresponding context words used in Experiment 1.

Target			Context
SW	PH	Non-PH	
BARN	BAHRN	MAHRN	FARM
BEEF	BEAF	GEAF	MEAT
BEER	BEERE	KEERE	WINE
BOMB	BAHM	SAHM	ATOM
BOOK	BOOCK	POOCK	NOVEL
BUTTER	BUDDER	SUDDER	BREAD
CHAIR	CHARE	THARE	TABLE
CLOUD	KLOWD	SLOWD	SKY
COAL	KOEL	THOEL	MINER
DANCE	DANSE	MANSE	PARTY
DARK	DARCK	VARCK	LIGHT
DESIRE	DEZIER	NEZIER	WISH
DIE	DIGH	KIGH	LIVE
DIRT	DURT	FURT	GROUND
FAST	PHAST	CHAST	SLOW
FEAR	FEER	VEER	AFRAID
FIRE	FYER	NYER	SMOKE
FIVE	PHIVE	THIVE	FOUR
FLOOR	FLOAR	PLOAR	CEILING
FORCE	PHORCE	THORCE	POWER
FOUND	PHOUND	KOUND	LOST
GIRL	GURL	SURL	BOY
GLASS	GLAS	PLAS	WINDOW
GOD	GAHD	HAHD	BIBLE
HAND	HANNED	DANNED	FOOT
HILL	HIL	ZIL	MOUNTAIN
HORN	HOARN	DOARN	FOG
HOT	HAHT	MAHT	COLD
LIES	LIZE	NIZE	TRUTH
LOW	LOE	KOE	HIGH
MINOR	MIGHNER	FIGHNER	MAJOR
MOST	MOAST	NOAST	LEAST
NAIL	KNALE	LALE	HAMMER
NAVY	NAYVY	LAYVY	ARMY
NIGHT	NITE	GITE	DAY
NINE	NIGHN	BIGHN	EIGHT
NOSE	KNOES	KOES	EYES
NUMBER	NAMBUR	DUMBUR	PHONE
NURSE	NERSE	DERSE	DOCTOR
PEACE	PEES	DEES	WAR
QUEEN	KWEEN	DWEEN	KING
RICH	RITCH	LITCH	POOR
ROAD	WRODE	TRODE	STREET
ROCK	ROK	TOK	STONE
ROUND	WROUND	LOUND	SQUARE

SAD	SADD	YADD	HAPPY
SAIL	SAYL	THAYL	BOAT
SAND	SANDE	MANDE	BEACH
SLEEP	SLEAP	CLEAP	BED
SMOOTH	SMUTHE	SLUTHE	ROUGH
SONG	SAWNG	NAWNG	MUSIC
SOUTH	SOUTHE	POUTHE	NORTH
STAR	STAHR	SLAHR	MOON
STORE	STOAR	SLOAR	MARKET
TAKE	TAICK	NAICK	GIVE
TALL	TAWL	NAWL	SHORT
UGLY	UGLIE	IGLIE	BEAUTIFUL
WEAK	WIEK	FIEK	STRONG
WOOL	WULL	TULL	SHEEP
WRONG	RONG	FONG	RIGHT

APPENDIX 2

Target source words (SW), pseudohomophones (PH), pseudohomographs (PG), and the corresponding context words used in Experiment 2.

Target			Context
SW	PH	PG	
AWAY	AWEIGH	AWELTH	GONE
BEEF	BEAF	BELF	ROAST
BIRD	BURD	BORD	HAWK
BUTTER	BUDDER	BUSHER	BREAD
CAKE	CAIK	CAWK	BAKE
CHAIR	CHARE	CHARK	TABLE
CLOSED	CLOZED	CLOOED	OPEN
CLOTHES	CLOZE	CLOME	WEAR
COUGH	COPH	COCH	SNEEZE
CRY	CRIE	CRIT	WEEP
DRY	DRIE	DROE	WET
FAR	PHARR	PHAER	NEAR
FIRE	FYER	FLER	SMOKE
FIVE	FYVE	FOVE	FOUR
FRAUD	FRAWD	FRAID	FAKE
FRUIT	FRUTE	FRUST	GRAPE
GOD	GAHD	GERD	BIBLE
GROAN	GRONE	GRONT	MOAN
HEAD	HEDD	HEND	ACHE
HOME	HOAM	HORM	HOUSE
HURT	HERT	HORT	PAIN
JUICE	JOOCE	JORCE	CIDER
KNIFE	NYFE	NOFE	BLADE
LANE	LAIN	LARN	ROAD
LATE	LAIT	LART	EARLY
LEAST	LEEST	LERST	MOST
LIES	LIZE	LIDE	TRUTH
LIGHT	LITE	LITH	GREEN
LOOSE	LUSE	LESE	TIGHT
LOVE	LUV	LIV	HATE
MINE	MYNE	MENE	YOUR
MUTE	MEWT	MERT	DEAF
NAIL	NALE	NARL	HAMMER
NAVY	NAYVY	NARVY	FLEET
NINE	NYNE	NUNE	EIGHT
PAPER	PAYPER	PARPER	FILE
PEACE	PEECE	PERCE	WAR
PEOPLE	PEEPLE	PERPLE	GROUP
PIE	PYE	POE	PIECE
PIPE	PYPE	PUPE	STOVE
RAIN	RANE	RAND	CLOUD
ROAR	RORE	ROTR	LION
ROOTED	RUTED	RETED	DEEP
RYE	RIE	REE	OATS
SAIL	SAYL	SARL	BOAT

SERVE	SURVE	SARVE	OBEY
SHEEP	SHEAP	SHEMP	HERD
SHOE	SHOOH	SHORN	SOLE
SLEEP	SLEAP	SLEMP	DREAM
SOAP	SOPE	SORP	SUDS
SOUP	SUPE	SUPL	EAT
TEETH	TEATH	TERTH	BITE
TILE	TIAL	TIRL	FLOOR
TOMB	TOOM	TORM	GRAVE
TRACKS	TRAX	TRAW	TRAIN
TREE	TREA	TREM	PINE
TWO	TWOO	TWOK	ONE
WEB	WHEB	WERB	SPIDER
WEIGHT	WATE	WRAT	TON
WHEAT	WHEET	WHEST	FLOUR
WHITE	WHYTE	WHOTE	CHALK
WRITE	WRYTE	WRUTE	READ
YOUNG	YUNG	YING	OLD
ZERO	ZEERO	ZEBRO	NONE

APPENDIX 3

Context source words (SW), pseudohomophones (PH) with their summed bigram frequencies, pseudohomographs (PG) with their summed bigram frequencies, and the corresponding target words used in Experiments 3 and 4. Stimuli marked † did not appear in Experiment 3.

Target			Context
SW	PH	PG	
ACHE	AIKE, 204	ARRE, 232	HEAD
BAKE	BAIK, 192	BAWK, 78 *	CAKE
BIBLE	BYBLE, 64	BOBLE, 83	GOD
BITE	BIGHT, 303	BIRET, 147*	TEETH
BLADE	BLAID, 129	BLARD, 199	KNIFE
BLUE	BLOO, 24	BLOM, 185	SKY
BOAT	BOTE, 129	BOTS, 115 *	SAIL
BRAVE	BRAIV, 139	BRARV, 201	COWARD
BREAD	BREDD, 127	BREND, 191	BUTTER
BREAK	BRAIK, 139	BRACK, 187	GLASS
BRIGHT	BRITE, 179	BRITH, 195	DULL †
CHEAT	CHEET, 364	CHENT, 367	HOAX
CIDER	SIDER, 332	MIDER, 339	APPLE
CODE	COAD, 114	COID, 128	SECRET
CRATE	CRAIT, 120	CRAST, 195	BOX
DATE	DAIT, 166	DAST, 347	FRIEND
DEAF	DEFF, 51	DELF, 170	MUTE
DEAL	DEEL, 226	DERL, 193 *	BARGAIN †
DOOR	DORE, 429	DORN, 227 *	KNOB
DREAM	DREEM, 129	DRERM, 244	SLEEP
EARLY	URLY, 63	ORLY, 63	LATE
EAST	EEST, 255	ERST, 216 *	WEST
EAT	EET, 96	ERT, 6 *	DRINK †
EIGHT	AIT, 55	DIT, 77	NINE
FAME	FAIM, 175	FAUM, 59 *	FORTUNE †
FATE	FAIT, 181	FANT, 207	DESTINY †
FLOOR	FLORE, 311	FLOTR, 141 *	TILE
FOUR	FOAR, 127	FOGR, 45 *	FIVE
FRUIT	FRUTE, 102	FRUST, 129	PEAR
GLOBE	GLOAB, 93	GLOOB, 121	WORLD †
GLUE	GLOO, 21	GLAM, 37	STICK
GONE	GAWN, 88	GERN, 174	AWAY
GOAT	GOTE, 121	GOOT, 233	MILK
GOLD	GOALD, 236	GOOLD, 233	JEWEL †
GRADE	GRAID, 146	GRALD, 178	SCHOOL
GRAVE	GRAIV, 131	GRARV, 193	TOMB
GRIEF	GREEF, 151	GROEF, 114 *	SORROW
GROUP	GRUPE, 57	GRUSP, 77	PEOPLE
HATE	HAIT, 321	HANT, 347	LOVE
HOPE	HOAP, 67	HORP, 177	DISPAIR
HOUSE	HOWSE, 217	HOLSE, 234	HOME
JAIL	JALE, 92	JALL, 232	PRISON †
LAKE	LAIK, 182	LASK, 160	POND †
LIGHT	LITE, 400	LITH, 596	DARK
MAIL	MAYL, 168	MAAL, 140 *	LETTER †

MAIN	MAYN, 168	MARN, 212	STREET
MEAL	MEEL, 213	MERL, 180 *	FOOD †
MOAN	MONE, 254	MOND, 271	GROAN
MONTH	MUNTH, 108	MINTH, 151	YEAR †
MOST	MOAST, 268	MOYST, 204 *	LEAST
NAVY	NAIVY, 53	NARVY, 75	MARINE †
NEAR	NEER, 298	NEBR, 37 *	FAR
NEAT	NEET, 238	NERT, 211 *	TIDY †
NONE	NUNN, 15	NANO, 113	ZERO
OATS	OTES, 46	ONTS, 177	RYE
OBEY	OBAY, 30	OBLY, 73	SERVE
ONE	WUN, 23	WEN, 34	TWO
PAIN	PAYN, 87	PARN, 131	HURT
PIECE	PEECE, 122	PRECE, 164	PIE
PHONE	FOAN, 129	FORN, 188	NUMBER †
RAIN	RANE, 184	RAWN, 90 *	MUD †
READ	WREED, 215	WREND, 148 *	WRITE
ROAD	ROED, 117	ROND, 243	HIGHWAY
ROOM	RUME, 242	REME, 245	BED †
SAFE	SAIF, 235	SARF, 189	SECURE †
SAME	SAIM, 237	SARM, 199 *	OTHER †
SHADE	SHAID, 150	SHALD, 282	TREE
SIX	SIKS, 107	SILE, 140	SEVEN †
SMOKE	SMOAK, 55	SMONK, 131	FIRE
SNOW	SNOE, 64	SNOP, 74	WINTER †
SOAP	SOPE, 145	SOPH, 118 *	BATH †
SOLE	SOAL, 137	SORL, 240	SHOE
SOUP	SUPE, 73	SURP, 74	SALAD †
STATE	STAIT, 223	STANT, 255	POLICE †
STEAK	STAIK, 213	STREK, 187 *	MEAT †
STONE	STOAN, 216	STORN, 214	ROCK †
STOVE	STOAV, 176	STORV, 208	PIPE
TABLE	TAYBLE, 162	TARBLE, 201	CHAIR
TAKE	TAIK, 165	TARK, 140 *	GIVE †
TAME	TAIM, 167	TARM, 129 *	WILD †
TAX	TAKS, 106	TAMS, 124	MONEY
TEACH	TEECH, 190	TERCH, 159 *	LEARN
TEASE	TEEZE, 58	TERLE, 133	ANNOY
THIEF	THEEF, 414	THREF, 391 *	STEAL †
TIGHT	TITE, 323	TITH, 519	LOOSE
TOAD	TODE, 130	TORD, 225	FROG
TON	TUNN, 22	TANN, 106	WEIGHT
TONE	TOAN, 152	TOON, 233	SOUND †
TRAIN	TRANE, 160	TRANK, 167	TRACKS
TROOP	TRUPE, 61	TRAPE, 105	ARMY
TRUTH	TROOTH, 94	TREETH, 61 *	LIES
WAR	WOAR, 123	WAYR, 112 *	PEACE
WADE	WAID, 256	WAAD, 105 *	WATER †
WAIT	WATE, 138	WATH, 334	STOP †
WHITE	WHYTE, 237	WHOTE, 326	BLACK
YOUR	YURE, 248	YURM, 38 *	MINE

---

*The summed bigram frequency of the pseudohomograph is lower than that of the corresponding pseudohomophone.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 231-242

The Contribution of Morphology to Word Recognition*

Laurie Beth Feldman†

Evidence of morphological processing was investigated in three word recognition tasks. In the first study, phonological ambiguity of the base morpheme in morphologically complex words of Serbo-Croatian was exploited in order to evaluate the claim that the base morpheme serves as the unit by which entries in the lexicon are accessed. An interaction of base morpheme ambiguity and affix characteristics was obtained and this outcome was interpreted as evidence that all morphological constituents of a word participate in lexical access. In the second study, facilitation due to morphological relatedness of prime and target was observed with Serbo-Croatian materials in the lexical decision and naming versions of the repetition priming task and results were interpreted as evidence of a morphological principle of organization among whole word forms in the lexicon. In the third study, morphological affixes of both English and Serbo-Croatian words were segmented from a source word and affixed to a target word more rapidly than phonemically matched controls. Results suggest that the morphological constituents of complex words are available in some word recognition tasks and that morphological knowledge is represented in the speaker's lexicon.

Morphology, the constituent structure of words, plays a focal role both in word processing and in sentence comprehension and production. It underlies the productivity of the word formation process and a word's fit into the syntactic frame of a sentence. In a number of languages, morphologically complex words consist of a root or base morpheme to which inflectional or derivational affixes are appended. Linguists distinguish between these two classes of formations. Words that differ in their derivational affixes but share a base morpheme (e.g., ACCOUNT, RECOUNT) are generally considered to be different lexical items and to have different meanings. Words that differ in their inflectional affixes (e.g., COUNTING, COUNTED) but share a base morpheme are generally considered to be versions of the same word, with the particular version that appears in a sentence being determined by the syntax of the sentence.

With regards to inflection, it is appealing to assume that forms are stored in terms of a base morpheme or stem and that particular inflectional variants are formed in the course of producing a sentence. The complexity of the formation rules and the unpredictability at the semantic level makes this option less plausible for derivations, however. Of course, the psychological utility of morphemes and the rules to combine them may, but need not, be consistent with the linguist's account. Morphologically complex words may be represented mentally either as whole forms or with reference to their constituents. These alternatives are not mutually exclusive (e.g., Caramazza, Laudanna, & Romani, 1988; Dell 1987) and they may apply differently to inflectional and derivational formations.

The experimental literature on word recognition focuses on three ways in which the morphological constituents of a word may be important. These include 1) morphologically-defined units for lexical access, 2) a morphological principle for organization (or representation) in the lexicon, and 3) sequencing errors in morphologically structured utterances. Models of word recognition differ with respect to the units by which a reader's knowledge

---

The research reported here was supported by National Institute Of Child Health and Development Grant HD-01994 to Haskins Laboratories.

about the semantic, morphemic, syntactic, phonological and orthographic properties of a word are accessed in the internal lexicon. One option, variously called the whole word, direct access, or addressed account, emphasizes the retrieval and match of full word units that are unanalyzed with respect to linguistic structure. A second option is called the phonological, indirect or assembled account. It posits that words are parsed into sublexical linguistic elements as part of the recognition process and that these units are used to search lexical memory. In addition to phonologically specified units such as those deemed critical for recognition of Serbo-Croatian words (Feldman & Turvey, 1983; Lukatela, Popadić, Ognjenović, & Turvey, 1980), candidates have included stems (i.e., base morpheme plus derivational affix) of morphologically complex words (Cole, Beauvillian, & Segui, 1989; Jarvella, Rob, Sandström, & Schreuder, 1987), an orthographically and morphologically defined first syllable or Basic Orthographic Syllabic Structure (i.e., BOSS) (Taft, 1979b, 1987), or a combination of whole word and morphemic elements (e.g., Caramazza, Laudanna, & Romani, 1988).

An understanding of principles of lexical organization derives principally from the lexical decision and similar tasks in which performance on a target word is improved by prior presentation of a related item or prime (Meyer, Schvaneveldt, & Ruddy, 1975; Stanners, Neiser, Hernon, & Hall, 1979). In the literature, both semantic/associative and morphological principles have been identified and the latter appear to be more robust. Morphological effects between bases and both derived and inflected forms as well as their inverse (Feldman & Fowler, 1987a) persist at lags of several intervening items in repetition priming. Moreover, changes in spelling or pronunciation do not diminish the effect significantly (Feldman & Fowler, 1987a; Feldman & Moskovljević, 1987; Fowler, Napps & Feldman, 1985; cf. Stanners et al., 1979) nor does the formal similarity of morphologically unrelated prime and target (e.g., pairs such as DIET and DIE) result in priming (Hanson & Wilkenfeld, 1985; Feldman & Moskovljević, 1987; Napps & Fowler, 1987). Finally, facilitation due to morphological relatedness as opposed to other types of semantic relatedness appears to entail distinct underlying mechanisms (Bentin & Feldman, 1990; Dannenbring & Briand, 1982; Feldman, 1990; Henderson, Wallis, & Knight, 1984). In short, the pattern of results observed with the repetition priming procedure has been interpreted as evidence that the morphological relations among words are represented in the lexicon. Although it is not without problems (see Bentin & Feldman, 1990), this outcome has been interpreted by this author in terms of a morphological principle of organization among full word forms in the lexicon.

Another lexical alternative advocated by Caramazza and his collaborators is that lexical representations are actually decomposed into morphological constituents (Caramazza et al., 1988). Most of their work was based on lexical decision rejection latencies with Italian materials and how they were affected by a nonword's component structure. For example, nonwords composed of morphological constituents in an illegal combination are difficult to process as evidenced by prolonged latencies in a lexical decision task. Similar results have been observed for Italian inflectional (Caramazza et al., 1988) and English derivational (Taft & Forster, 1975) morphology. Recently, however, morphological effects have been observed between pairs of Italian words in a double lexical decision task. Specifically, performance on word pairs that were formally similar in their initial morpheme but morphologically unrelated (i.e., the base morpheme in pairs such as PORTARE meaning to carry and PORTE meaning doors are both spelled PORT but represent different homophonous morphemes) was impaired relative to control pairs defined by the similarity of their initial (nonmorphemic) sequence of letters (i.e., pairs such as COLLO meaning neck / COLPO meaning a blow) (Laudanna, Badecker, & Caramazza, 1989). This outcome suggests processing interference among lexically distinct but homophonous base morphemes and has been interpreted as evidence for morphologically decomposed lexical entries.

Finally, speech error patterns (Cutler, 1980; Fromkin, 1973; Garrett, 1980; 1982; Stemberger, 1985) provide a source of compelling evidence for morphological constituents in the production of speech. In spontaneous speech, one prevalent type of error involves a reordering of morphemic elements. Either the stem or the affix of a word may migrate from the intended word to another site. Two patterns are relevant. First, word final elements that shift are typically morphemic elements as contrasted with phonemically equivalent but nonmorphemic segments. Second, inflectional suffixes are more vulnerable than derivational suffixes (Garrett, 1982). Collectively, these observations as well as the pattern observed in some clinical populations (e.g., Badecker & Caramazza, 1989) indicate that the constituent structure of morphologically complex words is available to the production mechanism and that inflectional and derivational forms may be represented differently (see also Miceli & Caramazza, 1988).

In summary, recent experimental work on the psychology of morphemes centers around effects that arise 1) preparatory to lexical access such as in the definition of units by which lexical representations can be accessed 2) in the lexicon itself, either as a principle of lexical organization or in the representational format of lexical entries 3) as a misordering of elements in an utterance. In the remainder of this chapter, three studies in Serbo-Croatian and a preliminary study in English that probe morphological effects in word recognition will be described and their relevance to issues of how morphology may contribute to word recognition will be discussed.

Interaction among morphological constituents

The first study attempts to assess the claim from English that morpheme-derived units provide the basis for lexical access (Caramazza et al., 1988; Taft & Forster, 1975; Taft, 1979; Taft, 1987). According to this account, morphologically complex words are decomposed into their constituents and stripped of their affixes so that the base morpheme is the unit to which a match in the lexicon is made. Typically, recognition latencies for words are governed by the time to parse and evaluate the lexical status of the stem and perhaps the time to assess the fit of the stem with its affix. Both base morpheme or stem and whole word frequency effects have been reported (Taft, 1979). Of course, it is possible that whole word effects actually reflect evaluation of the combination of base morpheme and affix combination.

The first study focuses on the phonological characteristics of the base morpheme and the experimental manipulation uses the phonological ambiguity effect (e.g., Feldman & Turvey, 1983; Lukatela et al., 1980) that is so robust in Serbo-Croatian. Specifically, it evaluates the claim that affixes are stripped from morphologically complex words prior to lexical access, that the base morpheme (in isolation or with any derivational affix) provides a unit by which lexical access can succeed and that the compatibility of affix and base is checked in a later operation (Taft, 1979a).

Critical to the phonological ambiguity manipulation is the fact that two alphabets, Roman and Cyrillic, are used interchangeably in parts of Yugoslavia. Due to the overlap between the two alphabet sets, some letter sequences can be assigned two distinct phonological interpretations, either (or both) of which can be meaningful. Moreover, all words in Serbo-Croatian can be transcribed in both alphabets and the alphabet sets are not generally mixed within a word or contiguous text. For example, the word meaning MARIGOLD is written NEVEN in Roman and HEBEH in Cyrillic. The former can be assigned only one phonological interpretation whereas the later is phonologically equivocal. In the experiments on ambiguity, the difference in recognition latencies in lexical decision (or naming) to phonologically ambiguous and unequivocal forms of the same word (i.e., HEBEH minus NEVEN) provide the measure of phonological ambiguity (see Turvey, Feldman, & Lukatela, 1984).

In contrast to most work on phonological ambiguity in Serbo-Croatian where ambiguous forms and their unequivocal controls were generally whole word units, in one study (Feldman, Kostić, Lukatela, & Turvey, 1983), the unit of phonological ambiguity was defined at the level of the morpheme. Experimental materials consisted of eight (four letter feminine or neuter) nouns whose base forms when transcribed in Cyrillic were phonologically bivalent but when transcribed in Roman were phonologically unequivocal. To each base morpheme was appended either a nominative suffix (viz., A or O) which was alphabetically (but not phonologically) bivalent in that it was pronounced and written the same way in Roman and in Cyrillic or a dative case affix (viz., I or U) which was neither alphabetically nor phonologically ambiguous.

Because in Serbo-Croatian, recognition latencies to a noun vary with inflected case such that nominatives tend to be faster than datives (Lukatela, Gligorijević, Kostić, & Turvey, 1980), the critical comparisons always held affix constant and focused on the difference between presenting forms of the base morpheme that were phonologically bivalent and phonologically unequivocal. That is, ambiguous and unequivocal forms of the same base morpheme were compared. Results indicated a significant effect of (phonologically ambiguous or unequivocal) base morpheme when the affix was phonologically ambiguous (viz., nominative) but not when it was phonologically unequivocal (viz., dative). Within the affix stripping decomposition account, differences among forms that share a base morpheme but differ with respect to affix may be accommodated by a check of affix-stem compatibility that occurs after the base morpheme has been segmented and accessed in the lexicon. However, differences that reflect phonological characteristics of the base morpheme are not anticipated. Moreover, in that study, when affixes were phonologically ambiguous, the difference between the ambiguous and unequivocal forms of the base morpheme increased as the number of ambiguous letters in the base morpheme increased.

As argued in that paper, the interaction of base morpheme ambiguity and affix can be interpreted within the decomposition before access framework if alphabet specification is retained after affixes are stripped. The effect of number of ambiguous letters which has been replicated in other experimental contexts (Feldman & Turvey, 1983; Lukatela, Feldman, Turvey, Carello, & Katz, 1989; Lukatela, Turvey, Feldman, Carello, & Katz, 1989) is more problematic, however. Therefore, the authors interpreted these data as evidence that both the base morpheme and the affix (either together or separately) contribute to lexical access time.

The first study presented here builds on these results and extends them to a set of twenty four verbs that possess a phonologically ambiguous base morpheme when written in one alphabet and a phonologically unequivocal analog when written in the other. Here, in contrast to that earlier work (Feldman et al., 1983), ambiguity of the base morpheme was equally distributed across alphabets. For example, the base morpheme of the verb meaning CALL is phonemically unique when transcribed in Roman JAV /jav/ but phonemically ambiguous when written in Cyrillic. (JAB can be pronounced as /jav/ which is meaningful or as /jab/ which is meaningless.) By contrast, the base morpheme of the verb meaning MELT is phonemically unique when written in Cyrillic TOП but phonemically ambiguous when written in Roman TOP (pronounced as /top/ or as /tor/). To the base morpheme was appended either an inflectional affix that was ambiguous with respect to alphabet (viz., E) or a verbal affix that included a letter that unequivocally specified alphabet in both its Roman and Cyrillic form (viz., IMO, ИMO). To summarize, verbal base morphemes were presented in their Roman and their Cyrillic transcriptions where one of those forms was phonologically ambiguous and one was phonologically unequivocal. Affixed to the base morpheme was 1) the inflectional affix for third person plural (i.e., E) which is alphabetically (but not phonologically) ambiguous because it is written the same way in Roman and in Cyrillic. 2) The first person plural affix which is written IMO in Roman and ИMO in Cyrillic. That is, it always includes a unique letter (viz., I or И) that necessarily specifies alphabet. Third and first person plural forms were selected because recognition latencies tend to be nearly equivalent in their unequivocal form (Kostić, personal communication). Skilled readers who were undergraduates at the University of Belgrade and were fluent in both alphabets served as subjects in either a lexical decision or a naming task. To anticipate, a simple decomposition account can accommodate effects of phonological ambiguity to the base morpheme but not an interaction of base morpheme and affix.

Latencies and errors to ambiguous base morphemes and their unequivocal controls for ambiguous and unequivocal affixes are summarized in Table 1. Effects are reported when analyses of variance treating both subjects and items as random variables were significant. For decision latencies (for times between 1400 ms and 400 ms), the effect of base morpheme ambiguity was significant as was the interaction of affix by base morpheme. For decision errors, a significant effect of base morpheme and affix as well as an interaction of base and affix was observed.

Table 1. *Decision latencies (and errors) to morphologically complex verb forms with phonologically ambiguous or unequivocal components.*

AFFIX:	AMBIGUOUS		UNIQUE	
STEM:	AMBIGUOUS	UNEQUIVOCAL	AMBIGUOUS	UNEQUIVOCAL
EXAMPLE:	JABE	JAVE	JABИMO	JAVIMO
LEXICAL DECISION:				
	729	671	677	664
	(28)	(14.3)	(9.2)	(6.6)
NAMING:				
	616	588	626	613
	(25.9)	(12.5)	(17.6)	(11.8)

The results of *t*-tests indicated that mean decision latencies for pairs such as JABE (729 ms) and JAVE (671 ms) that differ only with respect to ambiguity of the base morpheme but retain the same affix were significantly different $F1(1, 27)=18.01$, $MSe=4873$, $p<.001$. Pairs such as JAVIMO (664 ms) and JABИBO (677) that differed with respect to ambiguity of base morpheme[1] but shared the same affix (in different alphabet forms) did not differ. Mean errors for pairs such as JABE (28) and JAVE (14.3) were also significantly different $F1(1,27)=22.5$, $MSe=233$, $p<.001$. Finally, pairs such as JAVIMO (9.2) and JABИBO (6.6) that differed on ambiguity of base morpheme but shared the same affix (in different alphabet forms) did not differ.

In the analysis of naming latencies (for times between 1400 ms and 400 ms) only the effect of base morpheme ambiguity was significant. Naming errors, however, revealed a significant effect of base morpheme as well as an effect of affix and a marginally significant interaction of base and affix.

The results of *t*-tests indicated that mean naming latencies for pairs such as JABE (616 ms) and JAVE (588 ms) that differ only with respect to ambiguity of the base morpheme but retain the same affix were significantly different $F1(1,35)=4.72$, $MSe=5652$, $p<.04$. Pairs such as JAVIMO (613 ms) and JABИBO (626 ms) that differed with respect to ambiguity of base morpheme but shared the same affix (in different alphabet forms) did not differ significantly.

Mean errors in naming for pairs such as JABE (25.9) and JAVE (12.5) were significantly different $F1(1,35)=45.08$, $MSe=144$, $p<.001$. Pairs such as JAVIMO (17.6) and JABИBO (11.8) also differed $F1(1,35)=7.27$, $MSe=166$, $p<.01$. Pairs such as JAVE (12.5) and JAVIMO (11.8) did not.

If morphologically complex words were simply stripped of their affixes prior to lexical access then assuming that phonological codes are computed prelexically in Serbo-Croatian, one might expect to observe ambiguity effects, regardless of the phonological characteristics of the affix. If the compatibility of the affix with the stem also influenced latencies then one might obtain affix effects as well. Instead, in the present experiment, characteristics of the stem (i.e., stem ambiguity) and of the affix (i.e., alphabet specificity) interacted. This outcome replicates the results of Feldman, Kostić, Lukatela, and Turvey (1983) which focused on a small set of inflected nouns and it extends those findings to a more extensive set of inflected verb forms, some of which were ambiguous in their Roman transcription and some of which were ambiguous in their Cyrillic transcription. It also includes both lexical decision and naming as dependent measures. In contrast to the original work, analyses of number of ambiguous letters are not provided here. The general outcome is consistent with the claim that the base morpheme alone is not a sufficient unit for lexical access in Serbo-Croatian because alphabet information from the inflectional affix is preserved for lexical access. In conclusion, both morphological components must be analyzed in the course of word recognition.

Morphological effects in the lexicon

The second study in the series provides evidence that at least under some conditions, morphological effects do arise in the lexicon. The experimental strategy entails a comparison of the lexical decision and naming measures in order to insure that the facilitation due to morphological relatedness in a repetition priming task does not reflect task-specific processes. When analogous effects arise in lexical decision and naming, they are interpreted as revealing lexical processes. Effects observed in lexical decision without naming are sometimes treated as post-lexical in origin (Balota & Chumbley, 1984; 1985). However, others have claimed that when pseudowords are structured like words and subjects are required to process materials with respect to lexical knowledge, even though effects tend to be somewhat attenuated in naming relative to lexical decision both tasks reveal lexical effects (Paap, McDonald, Schvaneveldt, & Noel, 1987; see also Monsell, Doyle, & Haggard, 1989).

Significant facilitation among visually presented morphologically related words in the variant of the lexical decision task known as repetition priming (Stanners et al., 1979) is well documented and is generally interpreted as evidence that morphological relationships constitute a principle of organization within the internal lexicon. Typically, morphologically related primes reduce target decision latencies and error rates. Primes and targets can be unaffixed, inflected or derived forms in either the same or different modalities. Sometimes the reduction with morphological relatives as primes is equivalent to the effect of an identical repetition (Fowler et al. 1985). Effects of morphological relatedness persist in the lexical decision task across a variety of languages including Serbo-Croatian (Feldman & Fowler, 1987a) and Hebrew (Bentin & Feldman, in press) as well as English (Fowler et al., 1985; Feldman, in press) and American Sign Language (Hanson &

Feldman, 1989). It is unclear from these studies, however, whether these morphological effects are specific to the lexical decision task and reflect strategic or other post-lexical factors. In the second study, identical materials were presented to different subjects for either lexical decision or rapid naming.

Facilitation produced by two types of morphologically related primes, inflections and derivations were compared. In English, inflectional formations relative to derivational formations tend to be more similar in form and meaning to their base morpheme. Moreover, the number of inflectional affixes is severely limited relative to the number of derivational affixes. These limitations impede a rigorous experimental comparison between types of formations with English materials. By contrast, in Serbo-Croatian it is possible to identify inflection-derivation pairs with only minimal differences in form and meaning. One such contrast entails inflected verb forms that differ in aspect.

Aspect reflects temporal properties of the verb which may vary with particular verbs but generally include inceptive variants of stative verbs and iterative variants of verbs that describe discrete events (Bybee, 1985). It is relevant to note that the morphological status of aspectual formations is not universal across languages (compare Anderson, 1982 with Bybee, 1985). In the present study, it is assumed that aspect is a derivational formation for the class of verbs included because it was always the case that two distinct verbal entries existed in the dictionary. Each entry was formed around the same base morpheme but differed with respect to the set of inflectional affixes it required. If morphological relatedness by inflection and derivation are differently represented in the lexicon, then with controls for orthographic overlap, the magnitude of facilitation when forms that are inflectionally-related to the target are presented as primes should be enhanced relative to the pattern with derivationally-related primes.

Twenty seven word triples in Serbo-Croatian were selected. Each consisted of a target verb, an inflectionally-related prime and a derivationally-related prime. Targets consisted of present tense verb forms in the first or third person plural. Inflected forms were first person singular of those same verbs. Derived forms were first person singular of different verbs formed from the same base morpheme but differed in the temporal qualities of the action they conveyed.

Typically, the target and inflected prime were perfective forms and the derived prime was imperfective. Semantically, this distinction is relatively minor entailing contrasts between completed and durative actions such as HE SAT DOWN and HE WAS SITTING. Structurally, all members of a triple were composed of the same base morpheme but differed with respect to (the vowel of) the inflectional affix. For example, perfective forms of the base morpheme NOS meaning CARRY are formed around the vowel I such as NOSIM, NOSIS, NOSI...NOSE whereas imperfective forms are formed around A such as NOSAM, NOSAS, NOSA...NOSAJU. The orthographic and phonemic overlap of primes and their morphologically-related targets was matched by selecting third person plural forms ending in E as targets and necessarily as identity primes (e.g., NOSE), forms ending in IM (e.g., NOSIM) as inflectionally-related primes and verbs differing in aspect (e.g., NOSAM) as the derivationally-related primes. Inflectional and derivational primes were always presented in the same person and number.

Pseudoword triples were created by substituting vowels or consonants within base morphemes in order to create meaningless bases that were orthographically legal. To these, real inflected affixes were appended in order to create sets of pseudowords that differed only with respect to affix. As with words, pseudoword targets were preceded by identity, inflected and derivationally-related primes and inflected and derived forms differed only by the vowel in the verbal affix.

Three test orders were created. Each contained 114 items and included equal numbers of word and pseudoword targets preceded an average of ten items earlier in the list by a morphologically-related prime. In each test order, all three types of prime were equally represented. Across the three test orders, each word or pseudoword was preceded by all three types of morphologically related primes. The same test orders were presented to 27 subjects for a lexical decision judgement and to a different set of 27 subjects for a naming response. The presentation procedure was identical in both tasks and entailed presenting isolated words and pseudowords for 750 ms with an intertrial interval of 2000 ms.

Mean lexical decision and naming latencies for the second experiment are summarized in Table 2. An analysis of lexical decision latencies (less than 2 SD from the mean) and errors for words revealed a significant effect of type of prime Facilitation with inflectional primes was greater than with derivational primes $F(1,26)=8.11$, $MSe=1264$, $p<.008$. Neither pseudoword decision latencies nor errors revealed an effect of prime.

Table 2. *Facilitation in repetition priming with No Prime, Identity, Inflections and Derivations as primes.*

	TYPE OF PRIME				
	NO PRIME	IDENTITY	INFLECTIONAL	DERIVATIONAL	
	NOSE	NOSE NOSE	NOSIM NOSE	NOSAM NOSE	prime target
LEXICAL DECISION:					
	648	578	569	596	
	9.5	6.4	6.4	9.3	
NAMING:					
	597	569	567	582	
	10.3	9.5	7.4	11.52	

In word naming, latencies replicated the effect of type of prime. The differential effect of inflectional and derivation primes missed significance $F(1,26)=2.83$, $MSe=10046$, $p<.10$ as did the difference between no prime and derivational prime conditions. By contrast, the effect of inflectional primes was significantly greater than the effect of no prime $F(1,26)=12.09$, $MSe=1002$, $p<.002$. Finally, pseudoword naming latencies revealed a significant effect of type of prime that reflected differences between the no prime condition and each of the other conditions. The accuracy measure did not show the same pattern, however.

Comparisons of the magnitude of facilitation for targets preceded by inflectionally and derivationally related primes revealed significantly enhanced facilitation with inflections relative to derivations in lexical decision and a similar but dramatically attenuated pattern in naming. That the magnitude of priming in lexical decision is sensitive to type of morphological relation when formal properties of the affixes are controlled is important. It provides psychological support for a linguistic distinction and suggests that the failure to observe a similar effect in English may reflect differences in form between inflectional and derivational affixes. In naming, inflectionally related primes produced significant facilitation but the interpretation of derivational primes poses a problem. On one hand, they make targets no faster or more accurate than in the No Prime condition but on the other, they are not statistically slower than inflectional primes. Failure to obtain significant effects of type of prime may reflect generally diminished morphological facilitation in naming relative to lexical decision. An analogous effect of task has been observed under conditions of semantic facilitation (Katz & Feldman, 1983; Frost, Katz, & Bentin, 1987).

In summary, inflections do produce a systematic pattern of facilitation across tasks. This relation is represented in the lexicon and cannot easily be reduced to one of semantic activation. What remains unresolved is whether the relation among forms that share a base morpheme but differ with respect to derivational affixes is also represented in the lexicon. The results obtained with lexical decision suggest that they are. The results obtained in naming are equivocal in that the answer hinges on the interpretation of a weak pattern of facilitation. Because the pattern of means over task are very similar in that latencies to targets following derivations fall between latencies with no prime and those with inflectional primes, and because inflectional relations are represented in the lexicon it seems likely that the lexical status of inflections and derivations do differ. It is not obvious whether these data are better captured by a morphological principle of organization that differentiates inflections from derivations in terms of linkages among lexical entries or by decomposed lexical entries where inflectionally related forms are more tightly clustered than are derivationally related forms. What is certain is that facilitation among morphologically-related lexical entries cannot be reduced to patterns of semantic and orthographic activation.

Morphological segmentation

The final study in this report provides evidence that the morphological constituents of a word are sometimes available to a processing mechanism. The experimental manipulation exploits the fact that the morphemic structure of many words is not wholly transparent because the same sequence of letters can function morphemically in one context and nonmorphemically in another. For example, the final ER sequence in DRUMMER is morphemic whereas SUMMER contains only one morpheme. Similarly, the UN in UNTIE is morphemic whereas UNCLE is morphologically simple. Examples of morphologically complex and simple words are presented in Table 3.

Inspired by the pattern of errors in spontaneous speech, a new experimental technique was devised to obtain evidence for morphological constituents (Feldman & Fowler, 1987b). In the ***segment shifting task***, subjects are instructed to segment and shift a designated segment from a source word onto a target word and to name the new result aloud as rapidly as possible. In all cases, the product is a morphologically complex and real word. On some occasions, the affixed element (e.g., ER) is morphemic in status. Other times it is not. Morphemic and nonmorphemic endings are controlled for phonemic and syllabic structure (Tyler & Nagy, 1989). For example, subjects viewed source words such as SUMMER or DRUMMER for 500 ms then the ER was underlined, the target word PAINT appeared and a clock started. In each case, subjects said PAINTER and onset to vocalization was measured. The segment shifting procedure is depicted in Figure 1.

Table 3. *Examples of morphologically simple and complex words.*

ENGLISH:					
	PREFIX	BASE	SUFFIX: DERIVATION	SUFFIX: INFLECTION	MEANING
		DRUM	(M) ER		
		SOFT	EN		
		WRITE		S	
	RE	WRITE			
		WIN		(N)ING	
		BIT		(T)EN	
		SUMMER			
		OFTEN			
		WHITING			
		INNING			
		MITTEN			
SERBO-CROATIAN:					
		SLAV		OM	celebration
	DO	ČEK			reception
	S	PUST			descent
		CEV		I	pipes
		CED		I	he wrings
		KIŠ	ICA		drizzle
		TON		EM	I dive
		SLALOM			ski term
		DOBOŠ			drum
		MIŠICA			muscle
		FONEM			phoneme
		NOS		IM	I carry (perfective)
		NOS		AM	I carry (imperfective)

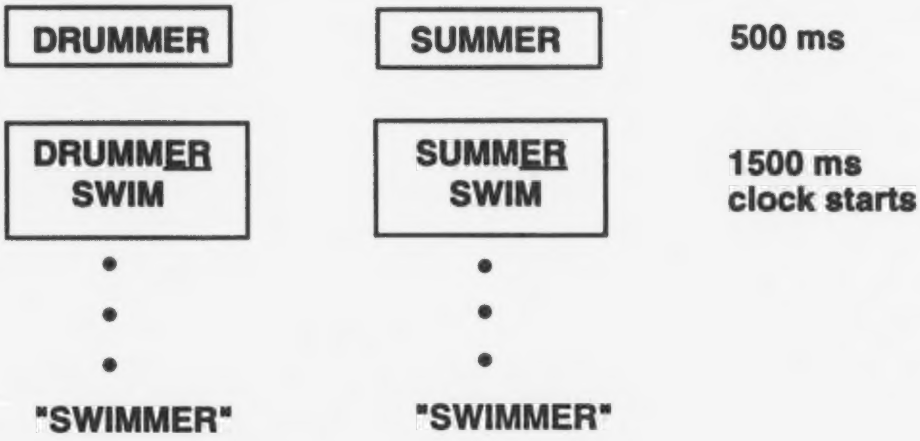


Figure 1. The segment shifting procedure.

In the first segment shifting experiment, thirty American college students were presented with forty pairs of English materials including both inflectional and derivational morphemic affixes or their nonmorphological controls. For example, experimental materials consisted of inflected words such as WINNING and its phonemically controlled nonmorphological pair INNING and derived words such as SOFTEN and its phonemically controlled nonmorphological pair OFTEN. Each subject saw only one member of a pair. The outcome was significantly faster (15 ms) naming latencies to the target derived from a source word with a morphemic ending than from one with a nonmorphemic ending. Due to the severe constraints on creating materials in the inflectionally impoverished language of English (all three inflectional affixes and six derivational affixes with no phonological influence on the base were included), no comparison between types of morphological formations was possible. Subsequent experiments in this series were conducted in Serbo-Croatian and explicit comparisons between types of morphological formations were performed.

The morphological effect in segment shifting was replicated with twenty six students at the University of Belgrade and sixteen pairs of Serbo-Croatian materials. Morphemic segments consisted of masculine singular instrumentals such as SLAVOM and nonmorphemic controls consisting of segments ending with the same sequence of letters without a morphemic function such as SLALOM. Morphemic segments were shifted (31 ms) significantly faster than their nonmorphemic controls. (Error rates were very low and produced no significant results.) Another condition of the experiment included homographic morphemes. Materials consisted of morphologically complex source words with morphological affixes that were compatible with the target word (same syntactic category and gender) and morphologically complex source words that were not. That is, the Serbo-Croatian analog of nominal or verbal S was shifted to another word of the same (consistent) or a different (inconsistent) word class. For example, the I from CEVI meaning PIPES in nominative plural or CEDI meaning HE WRINGS was shifted to the target word RAD in order to form the word RADI meaning HE WORKS. Note that in both source words the I is morphemic. What differs is the consistency or inconsistency of the syntactic category of the source word and the target word. In this condition, no significant effects of consistency (-8 ms) were observed. Shifting rates for I segments derived from verbal and nominal source words were equivalent.

In principle, the locus of the observed effect could be either the process of segmenting components from the source or affixing them to the target. Because there were no source-target compatibility effects, this outcome suggest the segmentation interpretation. Accordingly, morphemic affixes are more easily segmented from a source word than are nonmorphemic controls. Because morphemic and nonmorphemic affixes were controlled for phonemic structure, one interpretation of this outcome is that it reflects greater coherence of the morphemic letter sequences relative to the nonmorphemic sequences.

The linguistic productivity and lexical structure of inflectional but not derivational formations noted above, leads one to expect inflectional affixes to show enhanced segmentation effects relative to derivational affixes. The third experiment in this series investigated the effect of type of morphological formation on segment shifting. In this experiment, subjects had to substitute the affix from the source word for the affix on the target word and name the result aloud. Materials consisted of 24 Serbo-Croatian word pairs and the critical contrast compared the shifting latency of inflectional and derivational morphemic affixes with that of their nonmorphemic controls. Inflectional segments consisting of first person singular verbs ending in IM such as LEŽIM meaning I RECLINE and their nonmorphemic controls such as REŽIM meaning REGIME were shifted to inflected targets such as BACE meaning THEY THROW and subjects responded BACIM meaning I THROW. Likewise, derivational segments consisting of singular diminutives ending in ICA (and AK) such as BASTICA meaning LITTLE GARDEN and their controls such as KOŠTICA meaning SEED were shifted to target such as BUBA meaning BUG and subjects responded BUBICA meaning LITTLE BUG. In this experiment, the task required the subject to drop the original (morphemic) affix on the target (e.g., A) and to substitute the affix from the source word (e.g., ICA). Differences here are interesting insofar as they reveal processing contrasts between inflections and derivations that are controlled for phonemic structure and stress. It is important to note that this contrast of morphological types could not be performed in English due to the paucity of inflectional affixes and the stress pattern differences between inflections and many derivations.

Results in Serbo-Croatian revealed a significant morphological effect in segment shifting latencies for inflectional pairs (56 ms) but not for derivational pairs (-11 ms). No effects approached significance with the error measure. Results are consistent with the linguistic distinction between morphological types noted above and with the pattern of speech whereby inflections enter into speech errors more frequently than do derivations (Garrett, 1980). This finding suggests that the constituent structure of inflectional formations is more readily available to a processing mechanism than that of derivation.

CONCLUSION

It is evident from the data described above that the morphological components of a morphologically complex word can be accessed in the course of recognizing that word. Over a variety of experimental tasks, evidence of morphological analysis was observed. This process could not be reduced to one of identifying the stem within a complex form and using it as the unit by which representations in the lexicon were accessed, however, because phonological characteristics of the stem as well as the affix influenced recognition latencies under some conditions. Specifically, among Serbo-Croatian readers, phonological bivalence of a stem slowed decision and naming performance when the inflectional affix was ambiguous with respect to alphabet but not when it uniquely specified alphabet. If the stem in isolation provided the unit for lexical access then characteristics of the stem should influence all (or no) inflectionally related forms. Effects of affix as well as stem are consistent with the conclusion of Katz and his colleagues (Katz, Rexer, & Lukatela, 1990) that decomposition of morphologically complex forms does not occur prior to accessing the lexicon.

The pattern of facilitation among morphologically related but not among orthographically similar forms in repetition priming (Feldman & Moskovljević, 1987) in conjunction with the results of Experiment 2 provide compelling evidence that morphological relations are appreciated in the lexicon and that inflectional and derivational processing may differ even when semantic and orthographic similarity between these two types of formations are controlled. In both lexical decision and naming, inflectionally-related primes produced significant target facilitation at lags of 10 intervening items. Facilitation by derivationally-related primes in naming was weak relative to inflectional primes, however, and this outcome warrants further investigation. Without positing stems as the unit of lexical access, it appears that the presence of a common morphological component embedded in pairs of morphologically complex words can dramatically facilitate processing even when relatives are presented with lags of several items. Because the patterns emerge when the sound and spelling of the base morpheme are altered across prime and target as well as when they are preserved, this outcome has been interpreted by this author as reflecting a morphological principle of organization among whole word forms.

Results of the segment shifting task provide the most compelling evidence for morphological constituents in word recognition. When lexical and phonological output is controlled by experimental design, morphological segments are manipulated more efficiently than nonmorphological controls over a variety of morphological environments. Moreover, effects are more systematic for inflections than for derivations. This outcome

suggests that inflectional affixes are more easily segmented from the base than are derivational affixes and it is consistent with the linguists' claim that the internal structure of inflections are more accessible than that of derivations.

In summary, there is strong support from a variety of experimental paradigms for the psychological processing of the morpheme. The results in classical lexical decision and naming tasks do not convincingly differentiate between a morphological principle of organization in the lexicon and lexical representations that are decomposed into their morphological constituents, however. In a task that includes the manipulation of the inflectional affixes of a word and therefore resembles the task of inserting lexical entries into a syntactic frame of a sentence (Garrett, 1980; 1982), performance is constrained by the morphological status of the shifted component. Analogous effects with derivations are very weak or absent, however and this outcome supports a lexical distinction between inflection and derivation.

REFERENCES

Anderson, S. R. (1982). Where's morphology? *Linguistic Inquiry, 13,* 571-612.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decision a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340-357.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory & Language, 24,* 89-106.

Badecker, W. T., & Caramazza, A. (1989). A lexical distinction between inflection and derivation. *Linguistic Inquiry, 20* 106-116.

Bentin, S., & Feldman, L. B. (1990). Facilitation due to morphological and semantic relatedness at long and short lags: Evidence from Hebrew. *Quarterly Journal of Experimental Psychology,*

Bybee, J. L. (1985). *Morphology.* Amsterdam/Philadelphia: John Benjamins.

Caramazza, A , Laudanna,A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28,* 287-332.

Cole, P., Beauvillian, C., & Segui, J., (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory & Language, 28.* 1-13.

Cutler, A. (1980). Errors of stress and intonation. In V. A. Fromkin (Ed.), *Errors in linguistic performance.* New York: Academic Press.

Dannenbring, G. L., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology,* 36, 435-444.

Dell, G. S. (1987). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 285-321.

Feldman, L. B. (1991). Morphological relationships revealed through the repetition priming task. In M. Noonan, P. Downing, & S. Lima (Eds.), *Literacy and linguistics.* Amsterdam/Philadelphia: John Benjamins Pub. Co.

Feldman, L. B., & Fowler, C. A. (1987a). The inflected noun system in Serbo-Croatian: Lexical representation of morphological structure. *Memory & Cognition, 15,* 1-12.

Feldman, L. B., & Fowler, C. A. (1987b). Morphemic segments shift faster than nonmorphemic controls. Paper presented to the Psychonomic Society, Seattle, WA.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. *Psychological Research, 45,* 55-72.

Feldman, L. B., & Moskovljević, J. (1987). Repetition priming is not purely episodic in origin. Journal of Experimental Psychology: Learning, Memory & Cognition, 13, 573-581.

Feldman, L. B., & Turvey, M. T. (1983). Visual word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 288-298.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627-635.

Fowler, C. A., Nappe, S. E., & Feldman, L. B. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13,* 241-255.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 361-379.

Fromkin, V. A. (1973). *Speech errors as linguistic evidence.* The Hague: Mouton.

Frost, R., Katz, L. & Bentin, S. (1987). Strategies for visual word comparison and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 104-113.

Garrett, M. F. (1980). A perspective on research in language production. In E. C. Mehler, T. Walker, & M. F. Garrett (Eds.), *Perspectives on mental representation* (pp. 179-220). Hillsdale, NJ: Lawrence Erlbaum Associates.

Garrett, M. F. (1982). Production of speech. Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19-76). London: Academic Press.

Hanson, V. L., & Feldman, L. B. (1989). Language specificity in lexical organization: Evidence from deaf signers' lexical organization of ASL and English. *Memory & Cognition, 17,* 292-301.

Hanson, V. L., & Wilkenfeld, D. (1985). Morphophonology and lexical organization in deaf readers. *Language and Speech, 28,* 269-280.

Henderson, L. (1985). Toward a psychology of morphemes. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 15-72). London: Erlbaum.

Henderson, L., Wallis, J., & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. Bouhuis (Eds.), *Attention and performance* X. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jarvella, R. J., Rob, R., Sandstrom,, G., & Schreuder, R., (1987). Morphological constraints on word recognition. In A. Allport, D. Mackay, W. Prinz, & E. Scheerer (Eds.), *Language production and perception: Relations between speaking, listening, reading and writing.* London: Academic.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory & Cognition, 9,* 157-1661.

Katz, L, Rexer, K., & Lukatela, G. (1990). The processing of inflected words. *Psychological Research, 53,* 25-32.

Kostić, A. (1990). Processing of inflected morphology: Classic data reconsidered. *Psychological Research, 53,* 62-70.

Laudanna, A., Badecker, W., & Caramazza, A. (1989). Priming homographic stems. *Journal of Memory & Language, 28,* 531-546.

Lukatela, G., Popadič D., Ognjenović P., & Turvey, M.T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition, 8,* 124-132.

Lukatela, G., Feldman, L. B., Turvey, M. T., Carello, C., & Katz, L. (1989) Context effects in bi-alphabetical word perception. *Journal of Memory & Language, 28,*214-236.

Lukatela, G., Turvey, M. T., Feldman, L. B., Carello, C., & Katz, L. (1989). Alphabetic priming in bi-alphabetical word perception. *Journal of Memory & Language, 28.* 237-254.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word-recognition. In P. M. A. Rabbitt (Ed.), *Attention and performance V* (pp. 98-118). London: Academic Press.

Miceli, A. & Caramazza, A. (1988). Dissociation of inflectional and derivational processes. *Brain & Language,* 35, 24-65.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General,* 43-71.

Napps, S. E., & Fowler, C. A. (1987). Formal relationships among words and the organization of the mental lexicon. *Journal of Psycholinguistic Research, 16,* 257-272.

Paap, K. R., McDonald, J. E., Schvaneveldt, R. W., & Noel, R. W. (1987). Frequency and pronounceability in visually presented naming and lexical decision tasks. In M Coltheart (Ed.), *Attention and performance XII* (pp 221-243).

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18,* 399-412.

Stemberger, J. P. (1985). An interactive activation model of language production. *Progress in the psychology of language* (pp. 143-186). London: Lawrence Erlbaum Associates.

Taft, M. (1987). Morphographic Processing: The BOSS Re-emerges. In Max Coltheart (Ed.), *Attention and performance XII: The Psychology of Reading* (pp. 265-280). Hove: Lawrence Erlbaum Associates.

Taft, M. (1979a). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7,* 263-272.

Taft, M. (1979b). Lexical access via an orthographic code: The basic orthographic syllable structure (BOSS). *Journal of Verbal Learning and Verbal Behavior, 18,* 21-40.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14,* 638-647.

Turvey, M. T., Feldman, L. B., & Lukatela, G., (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and Reading.* (pp 81-90). London: Lawrence Erlbaum Associates.

Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory & Language, 28,* 649-667.

FOOTNOTES

[*] *Psychological Research, 53,* (1991).

[†] Also State University of New York at Albany.

[1] These forms also differed with respect to number of unique letters but (in contrast to Lukatela et al., 1990) this distinction did not significantly affect decision latencies.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 243-254

Lexical Mediation Between Sight and Sound in Speechreading[*]

Bruno H. Repp, Ram Frost,[†] and Elizabeth Zsiga[††]

In two experiments, we investigated whether simultaneous speechreading can influence the detection of speech in envelope-matched noise. Subjects attempted to detect the presence of a disyllabic utterance in noise while watching a speaker articulate a matching or a nonmatching utterance. Speech detection was not facilitated by an audio-visual match, which suggests that listeners relied on low-level auditory cues whose perception was immune to cross-modal top-down influences. However, when the stimuli were words (Experiment 1) there was a (predicted) relative shift in bias, suggesting that the masking noise itself was perceived as more speechlike when its envelope corresponded to the visual information. This bias shift was absent, however, with nonword materials (Experiment 2). These results, which resemble earlier findings obtained with orthographic visual input, indicate that the mapping from sight to sound is lexically mediated even when, as in the case of the articulatory-phonetic correspondence, the cross-modal relationship is nonarbitrary.

INTRODUCTION

The interaction of the visual and auditory modalities in word perception is of interest to psychologists concerned with the nature of the representation of words in the mental lexicon. That such an interaction exists has been demonstrated in many studies. For example, the popular cross-modal semantic priming paradigm (Swinney, Onifer, Prather, & Hirshkowitz, 1979) demonstrates facilitation of lexical access in one modality by the recent occurrence of a related word in the other modality. Visual articulatory information (i.e., a speaker's moving face) has long been known to aid the recognition of spoken words in noise (e.g., Erber, 1969; O'Neill, 1954), and, conversely, auditorily presented speech features which may not be intelligible by themselves can increase word recognition in speechreading (e.g., Breeuwer & Plomp, 1984, 1986). Cross-modal interactions can occur prior to word recognition: Printed single letters or nonword letter strings can facilitate the response to a phoneme presented in the auditory modality (Dijkstra, Schreuder, & Frauenfelder, 1989; Layer, Pastore, & Rettberg, 1990). Prelexical cross-modal influences have also been demonstrated when the visual information consists of articulatory gestures (McGurk & MacDonald, 1976): Simultaneous presentation of a spoken CV syllable and of a speaker's face uttering a different syllable can lead to the illusion of hearing the syllable suggested by the visual modality. This interaction even takes place prior to the categorization of the phonemes involved (Massaro & Cohen, 1990; Summerfield, 1987).

In a recent study, Frost, Repp, and Katz (1988) investigated whether influences from the visual modality can penetrate to earlier, precategorical levels of auditory perception by requiring their subjects to *detect* rather than recognize speech in noise. Auditory speech-plus-noise and noise-only trials were accompanied by a visual orthographic stimulus that either matched or did not match the masked speech. Frost et al. (1980) found that matching visual input did not improve subjects' speech detection performance, which suggested that the information subjects relied on (probably

---

This research was supported by NICHD Grant HD01994 to Haskins Laboratories. A brief report of the results was presented at the 31st Annual Meeting of the Psychonomic Society in New Orleans, LA, November 1990.

bursts of low-frequency spectral energy) was immune to cross-modal top-down influences. However, the visual input did have a strong effect on the bias parameter in this signal detection task: Subjects claimed to hear speech more often when they saw the word to be detected than when they saw a different printed word or no word at all. This *bias shift*, which may represent a genuine perceptual effect (viz., an illusion of hearing speech in noise), was evidently due to the fact that, in that study, the amplitude envelopes of the masking noises had been matched to those of the words to be masked. This so-called signal-correlated noise has very desirable properties as a masking agent (it enables precise specification of the signal-to-noise ratio and keeps that ratio constant as the signal changes over time) but it does retain some speechlike features. Although these features are not sufficient to cause perception of the noise as speech, let alone to identify a specific utterance, they do convey considerable prosodic and phonetic information. More specifically, the amplitude envelope conveys information about rate of speech (Gordon, 1988), number of syllables (Remez & Rubin, 1990), relative stress (Behne, 1990), and several major classes of consonant manner (Van Tasell, Soli, Kirby, & Widin, 1987). (See also Smith, Cutler, Butterfield, & Nimmo-Smith, 1989, who employed speech heavily masked by unmodulated noise.) Apparently, the subjects in the Frost et al. (1988) study automatically detected the correspondence between a printed word and an auditorily presented noise amplitude envelope. As a result, they perceived the masking noise as more speechlike and concluded that there was "speech in the noise." Frost et al. (1988) considered this an interesting and novel demonstration of rapid and automatic phonetic recoding in silent reading: Since signal-correlated noise is too impoverished to suggest a definite orthographic representation, the cross-modal correspondence must be established by mapping the print into an internal speechlike representation, specific enough to contain amplitude envelope features matching those of the noise and accessed rapidly enough to be linked to the transitory auditory stimulus.

According to many models of visual word recognition, the mapping from print to speech may be accomplished either via stored phonological codes attached to lexical entries or via prelexical spelling-to-sound conversion rules (see Patterson & Coltheart, 1987; Van Orden, Pennington, & Stone, 1990, for reviews). Hence it was especially interesting to find that the bias shift just described was reduced considerably when the materials were meaningless pseudowords (Frost et al., 1988: Exp. 2). Frost (1991) has replicated this finding in the Hebrew orthography, both with and without vowel symbols, using a within-subject design. His results suggest that the stronger bias shift for words than for nonwords is independent of spelling-to-sound regularity, and of the speed of processing the printed stimuli. It seems, therefore, that subjects' ability to detect the orthographic-acoustic correspondence in the speech detection paradigm is, at least in part, lexically mediated. That is, when the visual input is a word, it activates a lexical entry and, with it, an internal speechlike representation containing considerable phonetic detail, including amplitude envelope features. In contrast, when the visual input is a nonword, its internal phonetic representation (if any) must be assembled via analogy with known lexical items (Glushko, 1979) or via spelling-to-sound translation rules, and because of this piecemeal construction it may be less coherent or less vivid than the phonetic representation of a familiar word; hence the match with an auditory amplitude envelope is less evident.

Our aim in the present study was to further examine the hypothesis that detailed phonetic information is stored with, or is part of, lexical representations. We conducted two experiments analogous to Experiments 1 and 2 of Frost et al. (1988), but instead of print we employed a video recording of a speaker's face.

Visual articulatory information differs from orthography in several important ways. On one hand, whereas the relations of graphemic forms to phonologic structures are a cultural artifact, the relations of articulatory movements to phonological and phonetic structure are nonarbitrary. There is a natural isomorphism between visible articulatory movements and some acoustic properties of speech, particularly between the degree of mouth opening and overall amplitude. Therefore, lexical mediation may not be required for viewer-listeners to perceive a correspondence between the timing of opening/closing gestures and variations in signal amplitude.[1] On the other hand, visual articulatory information is less specific than print and generally conveys only distinctions among major consonant and vowel classes, the so-called visemes (see, e.g., Owens & Blazek, 1985). Visually observed speech gestures are often compatible with a number of lexical candidates. It may be hypothesized, therefore, that in order for a speechread utterance to be associated with the sound of a particular word, lexical access may be necessary, after all. Finally, we must note that articulatory information unfolds over time, whereas print is static and presents all information at once (provided it can be viewed in a single fixation). Thus there is an added dimension of temporal synchrony in audio-visual speech perception, which may enhance the interaction of the two modalities.

These considerations led us to hypothesize that the original finding of lexical mediation in the access of speechlike representations from orthography (Frost et al., 1988) might be replicated when the visual information consists of articulatory gestures: Subjects might be able to detect a correspondence between the speaker's gestures and auditory amplitude envelopes, but only when the stimuli are familiar words. In that case, the auditory envelope information would supplement the visual gestural information to constrain word identification.[2] A lexical representation would automatically link the two types of information, and a significant increase in perceptual bias on "matching" trials would be the result. However, when the speechread stimuli are clearly nonwords, lexical mediation would not occur, and this might also eliminate the bias shift, if it indeed originates at the lexical level.

Although the bias shift (i.e., the influence of visual information on perception of the *masking noise*) was of primary interest in our study, we also examined whether the detectability of the masked speech signal was influenced by seeing matching articulatory information. Our earlier studies with orthographic stimuli revealed absolutely no change in subjects' sensitivity to masked speech. However, because of the close relationship between visible articulatory information and speech acoustics, and because of the added dimension of audio-visual synchrony, we considered it possible that the speech gestures would aid listeners in separating the speech signal from the accompanying noise.

EXPERIMENT 1

Experiment 1 employed words as stimuli. Because we suspected that low-frequency energy provided the major cues for speech detection, and that utilization of these cues may be insensitive to cross-modal top-down influences, we included in Experiment 1 two auditory conditions, the first employing natural phonated speech and the second using whispered speech, which contains little low-frequency energy. These conditions provide very different opportunities for speechread information to exert an influence on auditory detection performance, as well as somewhat different amplitude envelopes for individual words to test the generality of the expected bias shift.

Method

Stimuli and design. The stimuli were 48 disyllabic English words with stress on the first syllable (examples: mountain, baby, canvas, etc.). A female speaker was recorded twice producing these words, once with normal phonation and list intonation, and once in a whisper, with the microphone much closer to her mouth. The first session was also videotaped, with the picture showing a frontal view of the speaker's face. Half the recorded words were used to generate the auditory stimuli. The same 24 words in each production mode (phonated and whispered) were digitized at 20 kHz and low-pass filtered at 9.6 kHz. Signal-correlated noise was generated from each word by a simple procedure that randomly reversed the polarity of half the digital sampling points (Schroeder, 1968). Such noise has exactly the same amplitude envelope as the original signal (obviously, since the envelope is derived from the rectified signal, i.e., regardless of the direction of the sound pressure change) but a flat spectrum, like white noise.[3] Speech-plus-noise stimuli were generated by adding the digital waveforms of each word and of its signal-correlated noise after multiplying them with weighting factors that added up to 1, so that the overall amplitude of the sum remained virtually the same. Two such weightings were used that, on the basis of pilot results, were expected to yield detection performance of 70-80 percent correct. In the phonated condition they corresponded to signal-to-noise (S/N) ratios of -12 and -14 dB. In the whispered condition, which was much more difficult, the S/N ratios used were -4 and -6 dB. All these ratios were well below the speech recognition threshold.

Within each production type (i.e., phonated or whispered) and S/N ratio condition, each of the 24 words appeared 6 times, 3 times as signal-plus-noise and 3 times as signal-correlated noise only. Each of these two auditory presentations occurred in three visual conditions: In the matching condition, the subjects saw the speaker produce the word that had been used to generate the auditory stimulus. In the nonmatching condition, they saw the speaker say a different disyllabic word, drawn from the 24 words not used as auditory stimuli. In the neutral condition, they saw the speaker's still face. The 6 audiovisual conditions for each of the original 24 words were distributed across 6 blocks of 24 trials according to a Latin square design. Thus each of the 24 words (in one of its two auditory incarnations) occurred exactly once in each block, and each of the 6 audiovisual conditions occurred 4 times per block (with different words). The 24 trials within each block were randomized. The more difficult condition with the lower S/N ratio always followed that with the higher S/N ratio, with the 144 trials of each following the same sequence. The phonated and whispered conditions also used the same stimulus sequences. The order of these production type conditions was counterbalanced across subjects.

The experimental video tapes were generated as follows. First, using professional video dubbing equipment, the video recordings from the phonated condition (with the original sound track on audio channel A) were copied one by one from the master tape onto the experimental tape, according to the randomized stimulus sequence for the video track. Each video segment started about 1 sec before, and ended about 1 sec after, the audible utterance. A view of the speaker's still face, of similar total duration, served as the neutral stimulus. About 3 s of black screen intervened between successive video segments. Second, the resulting audio track was digitized in portions, and the exact intervals between the onsets of the original spoken words were measured in displays of the digitized waveforms. (Most words began with stop consonants; for a few that began with nasals, the point of oral release following the nasal murmur was considered the onset.) Third, a computer output sequence was created containing the audio items to be substituted for the original utterances, according to the experimental stimulus schedule for the audio track, with exactly the same onset-to-onset intervals as those measured on audio channel A. Audio trials for the neutral condition were timed to start about 1 sec after the onset of the still face video. Finally, this auditory substitute sequence was output and recorded onto audio channel B, which was the one played back during the experiment.[4]

Subjects and procedure. The subjects were 12 paid volunteers, all native speakers of American English and claiming to have normal hearing. They were tested singly in a quiet room. The subject sat in front of a color monitor at a comfortable viewing distance and listened to the audio output over the video loudspeaker at a comfortable intensity. The task was described as one of speech detection in noise, and 24 practice trials using a higher signal-to-noise ratio (-10 dB in the phonated condition, 0 dB in the whispered condition) were provided without any accompanying video; these trials contained words not used as audio stimuli later on. Subjects were informed that a spoken word (either phonated or whispered, depending on the condition) was present in the noise on some of the trials. They were told to watch the video screen, but it was emphasized that what they saw had nothing to do with whether or not a word was hidden in the noise. The subjects wrote down their response (S for speech or N for noise only) on an answer form in the interval between trials. The whole experimental session (4 × 144 trials) lasted about 60 minutes.

Analysis. The data were analyzed in terms of the detectability and bias indices proposed by Luce (1963), which we call *d* and *b* here for simplicity, and which are comparable to the d′ and Beta indices of Signal Detection Theory. They are defined as

$$d = \ln[p(H)p(1\text{-}FA)/p(1\text{-}H)p(FA)]/2$$

and

$$b = \ln[p(H)p(FA)/p(1\text{-}H)p(1\text{-}FA)]/2$$

where p(H) and p(FA) are the proportions of hits and false alarms, respectively.[5] The *d* index (normally) assumes positive values similar to d′, and a positive *b* index indicates a bias to respond "S" (i.e., "speech present"). The indices we report below were computed for each subject and then averaged; however, we also computed indices for each item and did statistical analyses both ways. Separate analyses of variance were conducted on the phonated and whispered conditions, with S/N ratio and visual condition as within-subject factors; the F ratios for the subject and item analyses will be reported as F1 and F2, respectively.

Results

Detectability. In the phonated condition, the average d indices for the two S/N ratios were 2.14 (-12 dB) and 1.88 (-14 dB). This difference, which pitted the positive effect of practice against the negative effects of reducing the S/N ratio, was significant across items [F1(1,11) = 3.83, $p < .08$; F2(1,23) = 18.63, $p < .0004$], but is of little interest. The important result was that detection performance was unaffected by visual condition [F1(2,22) = 0.38, $p > .5$; F2(2,46) = 2.19, $p > .1$]; the average ratios in the three conditions were 1.95 (match), 2.02 (mismatch), and 2.07 (neutral). Thus, seeing a matching articulation did not aid speech detection. If anything, a match reduced sensitivity: In the item analysis, but not in the subject analysis, there was a significant S/N ratio by visual condition interaction [$F1(2,22) = 1.00$, $p > .3$; $F2(2,46) = 5.48$, $p < .008$]: At the higher S/N ratio, performance was best in the neutral condition and worst in the matching condition; this difference disappeared at the lower S/N ratio.

The average *d* indices were lower in the whispered than in the phonated condition, despite the much higher S/N ratios: 1.44 and 1.12, respectively, at the -4 dB and -6 dB ratios. The decline in sensitivity as a function of S/N ratio was signifiant [$F1(1,11) = 16.10$, $p < .003$; $F2(1,23) = 23.51$, $p < .0002$]. The performance levels were ideal for observing effects of visual condition. Still, there was no trace of a visual condition main effect [F1, $F2 < 1$]; the average values in the three conditions were 1.34 (match), 1.20 (mismatch), and 1.31 (neutral). Thus, even when low-frequency cues were eliminated, an audiovisual match did not facilitate detection performance. The S/N ratio by visual condition interaction was likewise nonsignificant [F1, $F2 < 1$].

Bias. We turn now to the results of primary interest. The *b* indices for both production mode conditions, averaged across the two S/N ratios, are shown in Figure 1 as a function of visual condition.

In the phonated condition, there was a strong bias to respond "S" in the matching condition, a lesser bias in the nonmatching condition, and hardly any bias in the neutral condition. This pattern of results matches that obtained with orthographic stimuli (Frost et al., 1988; Frost, 1991). The main effect of visual condition was highly significant [$F1(2,22) = 11.32$, $p = .0005$; $F2(2,46) = 29.79$, $p < .0001$]. Planned comparisons revealed reliable differences between the matching and nonmatching conditions [$F1(1,11) = 15.24$, $p < .003$; $F2(1,23) = 5.94$, $p < .03$], and between the nonmatching and neutral conditions in the item analysis [$F2(1,23) = 26.84$, $p < .0001$] but not in the subject analysis [$F1(1,11) = 3.70$, $p < .09$]. There were no significant effects involving S/N ratio.

In the whispered condition, the absolute b indices were much lower, but a very similar main effect of visual condition emerged [$F1(2,22) = 20.61$, $p < .0001$; $F2(2,46) = 55.14$, $p < .0001$]. There was a small bias to say "S" in the matching condition, no bias in the nonmatching condition, and a bias to say "N" in the neutral condition.

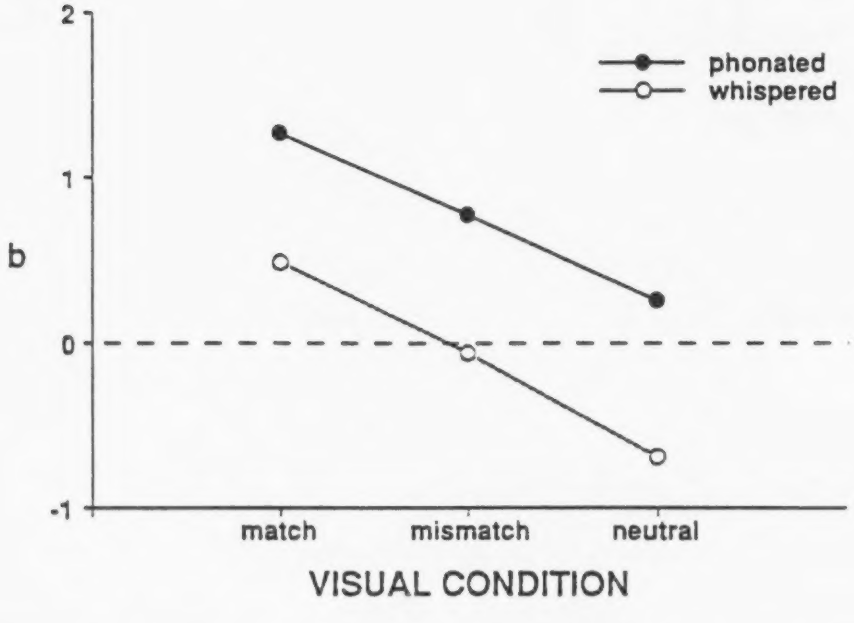


Figure 1. Bias indices in the phonated and whispered conditions of Experiment 1 (word materials) as a function of visual condition.

Planned comparisons showed reliable differences between the matching and nonmatching conditions [F1(1,11) = 14.42, $p < .004$; F2(1,23) = 20.08, $p < .0003$], and between the nonmatching and neutral conditions [F1(1,11) = 12.81, $p < .005$; F2(1,23) = 38.87, $p < .0001$]. There were no significant effects involving S/N ratio.

In summary, the results of Experiment 1 replicate almost exactly the findings of Frost et al. (1988) with orthographic word stimuli. Clearly, subjects were able to perceive a correspondence between speech gestures presented visually and amplitude envelopes presented auditorily. Like matching printed information, matching articulatory information, too, seems to create an illusion of hearing speech in amplitude-modulated noise. The bias shifts in the phonated and whispered conditions were equivalent. The difference between these conditions in absolute bias values must have a different origin (see General Discussion); whatever its cause, it is orthogonal to the relative bias shift that we are concerned with.

In order to determine whether the detection of correspondence between the speaker's articulatory gestures and the noise amplitude envelopes is lexically mediated, we examined in Experiment 2 whether nonword materials would produce the same effect.

EXPERIMENT 2

Since similar bias shifts were obtained in Experiment 1 regardless of production mode, only a phonated condition was employed in Experiment 2. Otherwise, except for the difference in materials, the experiment was an exact replication of Experiment 1. If there is a direct (i.e., prelexical) link between visible articulatory movements and the auditory amplitude envelope, then the results of Experiment 2 should replicate those of Experiment 1. If, on the other hand, this connection can only be established via the lexicon, then there should be no effect of audio-visual match on response bias. In particular, there should be no difference between the matching and nonmatching conditions; since it is conceivable that the mere presence versus absence of articulatory movements has an independent effect on response bias (see discussion below), the comparison with the neutral condition is less crucial.

Methods

The stimuli were 48 disyllabic nonwords stressed on the first syllable, produced by the same female speaker and videotaped. In Frost et al. (1988), orthographic nonwords had been generated from words by changing one or two letters. This would not do for speechreading because of the phonological ambiguity of visemes. To ensure that our stimuli were not speechread as English words, we used phonotactically atypical but easily pronounceable utterances containing the point vowels /a,i,u/ and visually distinctive consonants. (Examples: "vumuv," "kichaf," "fafiz," etc.). Twenty-four of the nonwords were used as auditory stimuli, the other 24 as nonmatching visual stimuli. The generation of stimulus tapes, the test sequences, and the procedure were identical with those in Experiment 1. Because detectability scores in the phonated condition of Experiment 1 had been somewhat high, the S/N ratios were set slightly lower in Experiment 2, at -13 and -16 dB. The subjects were 12 volunteers from the same general population. Two of them had participated in Experiment 1. They were informed that the utterances were meaningless.

Results

Detectability. The average d indices for the two S/N ratios were 1.62 and 1.11, respectively—significantly lower than the corresponding indices for phonated words in Experiment 1 [F1(1,22) = 5.84, $p < .03$; F2(1,46) = 10.96, $p < .002$, in a combined ANOVA], in part due to the somewhat lower S/N ratios used.[6] The main effect of S/N ratio was significant [F1(1,11) = 49.84, $p < .0001$; F2(1,23) = 24.88, $p < .0001$]. Surprisingly, there was also a significant main effect of visual condition here [F1(2,22) = 10.00, $p < .0009$; F2(2,46) = 6.50, $p < .004$]. This effect was due to a lower d index in the nonmatching condition (1.17) than in either the matching condition (1.51) or the neutral condition (1.42). In a combined ANOVA on the data of Experiment 1 (phonated condition) and of Experiment 2, with the added factor of lexical status (word/nonword), a significant interaction of visual condition and lexical status was obtained [F1(2,22) = 3.46, $p < .05$; F2(2,46) = 4.14, $p < .02$]. This suggests some inhibition or distraction caused by an audiovisual mismatch for nonwords, but no facilitation due to a match. The S/N ratio by visual condition interaction was nonsignificant.

Bias. The bias results are shown in Figure 2, averaged over the two S/N ratios. There was a significant effect of visual condition [F1(2,22) = 11.86, $p < .0004$; F2(2,46) = 13.68, $p < .0001$] but, as can be seen in the figure, it was entirely due to the matching and nonmatching conditions versus the neutral condition. There was absolutely no difference between the former two conditions, both of which exhibited a small positive bias. The effect of visual condition did not interact with S/N ratio.

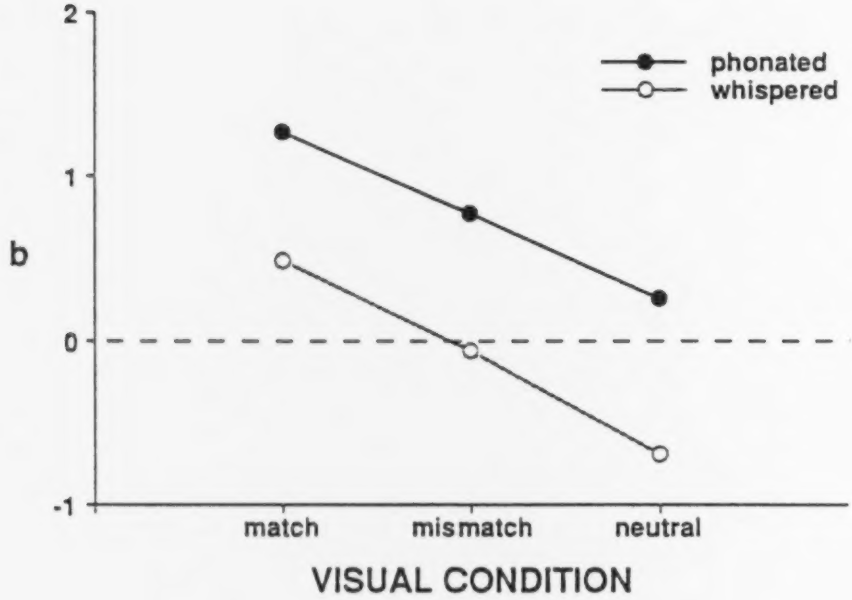


Figure 2. Bias indices in Experiment 2 (nonword materials) as a function of visual condition.

There was a marginally significant main effect of S/N ratio [F1(1,11) = 4.62, $p < .06$; F2(1,23) = 4.44, $p < .05$], due to an absolute decrease in the bias to say "S" when the S/N ratio was lowered. In order to compare directly the differences between the matching and the nonmatching conditions obtained for words and for nonwords, we combined in one ANOVA the data of Experiments 1 (phonated) and 2 for these two visual conditions. The interaction of visual condition and lexical status was significant across subjects [F1(1,22) = 11.11, $p < .004$] and nearly so across items [F2(1,46) = 3.54, $p < .07$].

GENERAL DISCUSSION

In the present study we examined the effect of a visual presentation of a speaker's face uttering words and nonwords on the detection of these words and nonwords in amplitude-modulated noise. Our experiments yielded three main results:

(1) There was no facilitative effect of audio-visual match on speech detectability.

(2) However, subjects recognized a correspondence between speech gestures and amplitude envelopes when the stimuli were words. Such an audio-visual match created an increased tendency to report the presence of speech in the detection task.

(3) This bias shift was absent when the stimuli were nonwords.

We will discuss these three results in turn.

Speech detection in noise and speechreading

The absence of a systematic effect of visual conditions on speech detectability is not too surprising in view of the fact that the task of detecting speech in noise requires only relatively low-level auditory processing. When the masking noise is coextensive with the speech and has the same amplitude envelope, as in our study, this means that the listeners must detect local spectral peaks that rise above the flat spectral level represented by the masking noise. When the speech is phonated, such peaks are most likely to occur in the lowest harmonics of voiced portions, and listeners therefore hear snippets of a human voice somewhere in the noise. Since speechreading does not provide information about the presence or absence of voicing, it cannot guide the listener to any portions of the signal that are especially likely to yield spectral evidence of voicing.

When the speech is whispered, listeners probably detect spectral prominences in the region of the second formant, or at higher frequencies if the word contains fricatives with strong noise components, such as /s/. This task is difficult because the speech itself has a noise source, and the S/N ratio must be raised considerably to achieve above-chance accuracy. Speechreading can provide some limited information about the occurrence of fricatives, but the most visible consonant articulations (/bpm/, /vf/, /θð/) have weak acoustic correlates, and fricatives such as /s/ were rare in our stimuli. Thus, there is not much to be gained from speechreading here either, and auditory detection strategies therefore seem to be uninfluenced by visual input.

There were two instances in which visual input did affect detectability scores, but the influence was negative rather than positive. In the phonated condition of Experiment 1, there was a tendency for detection to be best in the neutral condition, but only at the higher S/N ratio. More strikingly, in Experiment 2 detection scores were depressed in the nonmatching condition. Seeing articulatory movements may have had a slight distracting effect on listeners, especially when there was an obvious mismatch with the auditory input. Mismatches may have been more obvious in the nonword experiment, due to the different construction of the materials.

The bias shift for words

The result of primary interest is the relative change in bias as a consequence of audiovisual match. Our findings suggest that the visual presentation of speech gestures matching the auditory amplitude envelope causes an auditory illusion of hearing speech, similar to the illusion obtained by Frost et al. (1988) with printed stimuli. This may not seem surprising: If subjects can detect the correspondence between the auditory amplitude envelope and print, whose relationship to each other is merely conventional, then they certainly should also detect the correspondence between the envelope and articulatory movements, which are intrinsically linked. In particular, the visible time course of jaw opening is a direct optic correlate of the gross amplitude envelope. It is not necessary to invoke lexical access to explain the results for words. Lexical access probably did occur, however, due to the joint constraints effected by the auditory amplitude envelope and the visual articulatory information, and it probably happened more often in the matching than in the nonmatching condition.

Two aspects of subjects' sensitivity to audiovisual matches deserve comment. First, an effect of match was obtained even though the auditory and visual inputs were not in perfect synchrony; this suggests, in accordance with earlier findings (see Footnote 4), that temporal offsets smaller (and occasionally larger) than 100 ms do not interfere substantially with the detection of audiovisual correspondence, especially if the sound lags behind. Second, the bias shift was obtained for both phonated and whispered speech, even though the amplitude envelope of a given word was different in the two production modes. Since the same video was used in both conditions and relative bias shifts of the same magnitude were obtained, this means that the audio-visual match was equally good for both kinds of amplitude envelopes. The amplitude envelopes thus must have retained crucial phonetic properties across the change in phonation type (cf. Tartter, 1989, on phonetic information in whispered speech). The extent to which the speech amplitude envelope conveys invariant phonetic features is a worthwhile topic for investigation that has received only very limited attention so far (e.g., Mack & Blumstein, 1983; Nittrouer & Studdert-Kennedy, 1985).

There was one difference, however, between the phonated and whispered conditions: The absolute bias indices were considerably lower in the whispered condition. Since the masking noises were rather similar in the two conditions, the difference in bias must reflect differences in subjects' expectations of hearing speech. The greater difficulty of the whispered condition and the atypicality of whispered speech may have been sufficient reasons for subjects' relative conservatism, as reflected in the absolute bias indices.

So far, we have focused on the difference between the matching and nonmatching conditions for words, which constitutes the predicted bias shift. However, there was also a reliable difference between the nonmatching and neutral conditions, with the bias to say "S" being relatively greater in the nonmatching condition. This difference was also obtained in the earlier study with print (Frost et al., 1988). There are two possible interpretations: (1) The effect may represent a different kind of response bias, caused by *any* structured visual input (print or articulation) regardless of match. According to this view, there are really two bias shifts: a less interesting one (postperceptual response bias) that accounts for the difference between neutral and non-neutral conditions, and a more interesting one (perceptual in origin) that accounts for the difference between the matching and nonmatching conditions. (2) Alternatively, the difference between the neutral and nonmatching conditions may represent an effect of partial match. After all, the nonmatching stimuli had the same general prosodic pattern as the matching stimuli (i.e., two syllables with stress on the first). This may have been sufficient to obtain a small bias shift. According to this view, there is a single bias shift effect which is present in varying degrees in the matching and nonmatching conditions; the "nonmatching" condition really should have been called "partially matching" in that case.

The present data for word stimuli cannot decide between these two alternatives. However, a previous experiment that bears on the issue is Experiment 3 of Frost et al. (1988), which used orthographic visual stimuli. In that experiment, white noise without amplitude modulation was used as a masker. Thus, there was no auditory basis for either whole or partial matches. Yet, a difference in bias was obtained between the neutral condition and the other two conditions. This suggests that the first explanation given above is correct, at least for print.

The absence of a bias shift for nonwords

This suggestion seems to be confirmed by the present results: The difference in bias between the matching and nonmatching conditions, obtained for word stimuli in Experiment 1, was absent for nonword stimuli in Experiment 2. There was, however, a reliable difference between the neutral condition and the other two visual conditions even for nonwords, and this difference was similar in magnitude to that between the neutral and nonmatching conditions for words. If the relative bias increase in the nonmatching condition represented an effect of partial match, then it would be difficult to explain why an additional effect of complete match was obtained for words only. Therefore, the difference between the neutral and nonmatching conditions may well represent an "uninteresting" response bias, due to the occurrence of any verbal event in the visual modality.

However, the partial match explanation can still be upheld by noting that the partial match reflects only general prosodic characteristics (number of syllables, stress pattern) whereas the complete match reflects the added effect of matching segmental envelope characteristics as well as prosodic detail such as the exact timing pattern. To account for the effect of lexical status, one is then led to the interesting (but highly speculative) conclusion that the detection of segmental (and exact prosodic) cross-modal matches requires lexical access, whereas the detection of gross prosodic matches can occur without the involvement of the lexicon.

A similar conclusion was reached independently, and on the basis of quite different kinds of evidence, by Cutler (1986): In a cross-modal priming task, auditorily presented words drawn from semantically distinct pairs that differed only in stress pattern but not in segmental structure (quite rare in English; e.g., FORbear—forBEAR) had equal priming effects on lexical decision for visual targets that were semantically related to one or the other member of the pair. In other words, the auditory stress pattern did not constrain lexical access and only postlexically disambiguated the two semantic alternatives. Our results are complementary to those of Cutler in that they suggest that global prosodic information, including stress pattern, is processed independently of lexical access. This result makes sense when we consider the fact that prosodic parameters are not specific to speech but also play an important role in music, in animal communication, and even in environmental sounds. Lexical access, perhaps necessarily, is governed by speech-specific (segmental and detailed prosodic) properties of the acoustic signal; global prosodic properties, on the other hand, feed into the nonverbal systems of auditory event perception and emotion. They may also be processed in different parts of the brain.

The above interpretation remains speculative because we do not know what would happen on trials on which there is a striking *prosodic mismatch* between the auditory and visual inputs. An experiment including such trials remains to be conducted. Our results show very clearly, however, that an audio-visual match of segmental (and detailed prosodic) characteristics leads to a bias shift only for words, not for nonwords. This results replicates earlier findings obtained with print (Frost et al., 1988; Frost, 1991) and in fact is more dramatic: Whereas a small difference between matching and nonmatching conditions was consistently obtained with printed nonwords, there was no difference at all with speechread nonwords, perhaps because the latter were less similar to English words than the printed nonword stimuli. In the case of print, the results suggested that lexical access through the visual modality results in a detailed phonetic representation that shares amplitude envelope features with a matching signal-correlated noise. The alternative process of letter-to-sound translation by rule or analogy, which—according to traditional dual-route models—must be employed for nonwords, is either too slow to enable subjects to relate its product to the auditory stimulus or, more likely, does not result in a detailed, complete, or coherent phonetic representation. The latter interpretation is favored by Frost's (1991) recent results, which show that manipulations known to affect speed of word recognition (viz., word frequency and Hebrew vowel diacritics) have no effect on the magnitude of the bias shift for words; by implication, the absence of a bias shift for nonwords is probably not due to a slower processing speed. Can the same arguments be made in the case of speechreading?

In the introduction, we pointed out three important differences between print and visual articulatory information. Two specific aspects of speechreading, the temporal nature of the information and its nonarbitrary relation to the sounds of speech (including the amplitude envelope), led to the expectation that an effect of audio-visual match might be obtained regardless of lexical status. This was clearly not the case; thus, speechread information is not directly translated into a phonetic representation. The reason for this lies probably in the third aspect: The visual information is not specific enough. Inner speech consists of the sounds of words, not just of their amplitude envelopes, which are features of the complete sound patterns. Speechread information rarely specifies a unique word, however, and hence it does not (or only rarely) lead to lexical access in the case of isolated words, nor does it enable a viewer to construct a detailed sound pattern by direct translation, bypassing the lexicon. Normally, the incomplete information needs to be supplemented by additional information that constrains the possible lexical choices. The auditorily presented amplitude envelope probably functioned as such a source of supplementary information (see Footnote 2). In addition, its spectral masking power may have created the auditory illusion of hearing speech, as in the phonemic restoration effect (cf. Warren, 1984).

This role of the auditory amplitude envelope in conjunction with speechreading is somewhat different from the role Frost et al. (1988) attributed to it in their studies with print, where they saw it as *probing* into the process of lexical access from (unambiguous) print. In the case of speechreading, the noise envelope is not so much a probe as an active ingredient in the processes leading to lexical access. (When printed stimuli are made ambiguous, as in a recent, still unpublished study by Frost, the same is probably true.) The best way, then, to characterize what happened in our present experiments is that amplitude envelope information and speechread information often converged onto a lexical entry in the case of words, but failed to do so in the case of nonwords. Whether the bias shift for words was a direct consequence of this lexical convergence, or whether a separate postlexical process detected the match between the phonetic representation stored in the lexicon and the noise envelope, is a moot and probably unresolvable question. It may be concluded, however, that it is the lexically mediated activation of an internal phonetic representation that accounts for the illusion of hearing speech in the noise, and hence for the bias shift.

REFERENCES

Behne, D. M. (1990). The position of the amplitude peak as an acoustic correlate of stress in English and French. *Journal of the Acoustical Society of America, 8&7*, S65-66 (A)

Blamey, P. J., Martin, L. F. A., & Clark, G. M. (1985). A comparison of three speech coding strategies using an acoustic model of a cochlear implant. *Journal of the Acoustical Society of America, 77*, 209-217.

Breeuwer, M., & Plomp, R. (1984). Speechreading supplemented with frequency-selective sound-pressure information. *Journal of the Acoustical Society of America, 76*, 686-691.

Breeuwer, M., & Plomp, R. (1986). Speechreading supplemented with auditorily presented speech parameters. *Journal of the Acoustical Society of America, 79*, 481-499.

Cutler, A. (1986). *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*, 201-220.

Dijkstra, T., Schreuder, R., & Frauenfelder, U. H. (1989). Grapheme context effects on phonemic processing. *Language and Speech*, 32, 89-108.

Dixon, N. F., & Spitz, L. (1980). The detection of audiovisual desynchrony. *Perception, 9*, 719-721.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12*, 423-425.

Frost, R. (1991). Phonetic recoding of print and its effect on the detection of concurrent speech in noise. Manuscript submitted for publication.

Frost, R., Repp, B. H., & Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory and Language, 27*, 741-755.

Glushko, R. J. (1979). The organization of activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 674-691.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics, 43*, 137-146.

Grant, K. W., Ardell, L. H., Kuhl, P. K., & Sparks, D. W. (1985). The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *Journal of the Acoustical Society of America, 77*, 671-677.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138-1141.

Layer, J. K., Pastore, R. E., & Rettberg, E. (1990). The influence of orthographic information on the identification of an auditory speech event. *Journal of the Acoustical Society of America, 87*, Suppl. 1, S125. (A)

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. New York: Wiley.

Mack, M., & Blumstein, S. E. (1983). Further evidence of acoustic invariance in speech production: The stop-glide contrast. *Journal of the Acoustical Society of America, 73*, 1739-1750.

Massaro, D. W., & Cohen, M. M. (1990). Perception of synthesized audible and visible speech. *Psychological Science, 1*, 55-63.

McGurk, H., & MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America, 77*, 676-685.

Nittrouer, S., & Studdert-Kennedy, M. (1985). The stop-glide distinction: Acoustic analysis and perceptual effect of variation in syllable amplitude envelope for initial /b/ and /w/. *Journal of the Acoustical Society of America, 80*, 1026-1029.

O'Neill, J. J. (1954). Contributions of the visual components of oral symbols to speech comprehension. *Journal of Speech and Hearing Disorders, 19*, 429-439.

Owens, E., & Blazek, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research, 28*, 381-393.

Patterson, K., & Coltheart, V. (1987). Phonological processes in reading: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 421-447). Hove, East Sussex, UK: Lawrence Erlbaum Associates.

Remez, R. E., & Rubin, P. E. (1990). On the perception of speech from time-varying acoustic information: Contributions of amplitude variation. *Perception & Psychophysics, 48*, 313-325.

Repp, B. H., & Frost, R. (1988). Detectability of words and nonwords in two kinds of noise. *Journal of the Acoustical Society of America, 84*, 1929-1932.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 43*, 1735-1736.

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research, 32*, 912-920.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Swinney, D. A., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition, 7*, 159-165.

Tartter, V. C. (1989). What's in a whisper? *Journal of the Acoustical Society of America, 86*, 1678-1683.

Tillmann, H. G., Pompino-Marschall, B., & Porzig, U. (1984). Zum Einfluss visuell dargebotener Sprechbewegungen auf die Wahrnehmung der akustisch kodierten Artikulation. *Forschungsberichte des Instituts fur Phonetik und sprachliche Kommunikation* (University of Munich, FRG), *19*, 318-336.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97*, 488-522.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America*, 82, 1152-1161.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. Psychological Bulletin, 96, 371-383.

FOOTNOTES

[*] *Quarterly Journal of Experimental Psychology*, in press.

[†] Hebrew University of Jerusalem.

[††] Also Yale University.

[1] Kuhl and Meltzoff (1982) have shown that infants 18-20 weeks old perceive the correspondence between visually presented /i/ and /a/ articulations and the corresponding speech sounds. However, the infants did not recognize any relationship when the amplitude envelopes of these vowels were imposed on a pure tone, so they probably relied on spectral rather than amplitude information when listening to speech.

[2] It is known from research on possible aids for the hearing-impaired that the auditory speech amplitude envelope, even when carried just on a single pure tone, constitutes an effective supplement to speechreading (Blamey, Martin, & Clark, 1985; Breeuwer & Plomp, 1984, 1986; Grant, Ardell, Kuhl, & Sparks, 1985). Note that, in our experiments, it does not matter whether the word recognized is "correct" (i.e., the one intended by the speaker) or not, as long as it fits both the auditory and the visual information.

[3] Although the noise had a flat spectrum in its digital form, it was output through hardware designed to remove high-frequency pre-emphasis and thus had a sloping spectrum in its acoustic form. For the purposes of the present experiments, this was irrelevant.

[4] As we did not have equipment available to trigger the output sequence precisely and thus to ensure exact audio-visual synchrony, we started and restarted the output sequence manually until it seemed in synchrony with the video channel. Subsequently, we measured the onset asynchrony between audio channels A and B on matching trials, using two-channel digitization and digital waveform displays. If any asynchrony exceeded ±100 ms, we re-recorded the output sequence. Asynchronies within this range are difficult to detect (Dixon & Spitz, 1980; McGrath & Summerfield, 1985) and seem to have only a negligible effect on audiovisual speech perception (McGrath & Summerfield, 1985; Tillmann, Pompino-Marschall, & Porzig, 1984). Although we believed at the time to have satisfied this criterion, postexperimental checks revealed some inaccuracies in the test sequence specifications that led to onset asynchronies in excess of 100 ms for some stimulus combinations. These asynchronies were always such that the sound lagged behind the visual stimulus, which is less detectable than the opposite (Dixon & Spitz, 1980), and they occurred only in the phonated condition. Although this aspect should have been under better control, we have no indication that audio-visual asynchrony had any effect whatsoever on our results; in particular, as will be seen, the phonated and whispered conditions yielded very similar bias shifts.

[5] Values of 1/2N and 1-1/2N were substituted for proportions of 0 and 1, respectively. Due to the different frequencies of these substitutions and the nonlinear nature of the indices, the average *d* and *b* indices were not identical when computed across subjects and across items.

[6] Performance for nonwords was somewhat lower than expected on these grounds alone. Of course, this could have reflected a random difference between subject samples. However, Frost et al. (1988), too, found lower detection performance for nonwords than for words in different experiments, even though the words and nonwords were equally detectable when presented randomly within the same experiment (Repp & Frost, 1988; Frost et al., 1988: Exp. 3). It is as if subjects listened less carefully when they are presented with nonsense.

Haskins Laboratories Status Report on Speech Research
1991, SR-107/108, 255-262

Leap of Faith[*]
A review of Derek Bickerton's *Language and Species*

Michael Studdert-Kennedy[†]

> ...believing, as I do..., that the possession of articulate speech is the grand distinctive character of man..., I find it very easy to comprehend that some...inconspicuous structural difference may have been the primary cause of the immeasurable and practically infinite divergence of the Human form from the Simian stirps.
>
> —T. H. Huxley, *Man's Place in Nature* (1863/1915, p. 143).

Here is a book after Huxley's heart, arguing that language is indeed the distinctive character of *Homo sapiens*, the foundation of the species' self-reflexive consciousness and, in the end, of its control over the terrestrial environment. The book is the first of its kind, a closely reasoned attempt to fit linguistic theory into the framework of evolutionary biology. Actually, Bickerton has little to say about "articulate speech" (of this, more below), and the linguistic theory he is trying to square with biology is the current version of generative grammar. But since no other description of language has anything like the scope and detail of generative grammar, Bickerton's decision to follow Chomsky, "...arguably the Newton of our field" (p. 5), was perhaps a necessary step toward confronting the full complexity of language. The book is remarkable for its determined, if ultimately unsuccessful, endeavor to do justice to both evolutionary and linguistic theory, a task from which (as Bickerton remarks) professional linguists, despite lip service to biology, have resolutely flinched.

Of course, Bickerton himself is a professional linguist, a student of pidgins and creoles, and this fact has determined the highly original cast of his argument. But he has read widely in many fields including anthropology, evolutionary biology, philosophy, psychology, and neuropsychology, and this fact gives the argument its depth and scope. Although there are many passages of somewhat technical discussion and although the book demands constant attention, if you don't want to miss any tricks, Bickerton addresses a general audience in a brisk conversational style.

The gist of the argument is this. Anyone who accepts the theory of evolution must also accept that language, however complex, is an adaptation that evolved out of some prior system. Yet the gap between language and the most sophisticated systems of non-human animal communication is so vast that there seems to be no conceivable way of getting from there to here. This is the Paradox of Continuity.

The source of the paradox lies in a qualitative difference between the modes of human and non-human communication. So far as we know, all non-human systems address a limited set of topics essential for survival and reproduction (food, predators, sex). Humans, by contrast, can talk (at least in principle) about anything they choose. The difference is due not simply to differences in cognitive capacity (which, to a large degree, may be a consequence rather than a cause of language), but rather to formal differences in the mechanisms of communication. Animal calls and signs are structurally and functionally holistic: each unitary call refers to a whole situation. A spoken utterance, on the other hand, typically breaks the situation to which it refers into parts (agent, action, goal, instrument, locus, etc.). This can be done because the structure of language is itself componential: an effectively unlimited set of words is formed from a small set of sounds, and an infinite number of sentences can be formed from the words. This difference is the grounds of the "practically infinite divergence" in the communicative capacities of human and non-human animals. How did evolution bridge the gap?

To resolve the Continuity Paradox, Bickerton argues, we must recognize that language is primarily not a means of communication, but "...a system of representation, a means for sorting and manipulating the plethora of information that deluges us throughout our waking life" (p. 5). Before we can talk, we have to have something to say, and language is, in the first instance, a means of representing to ourselves what we have to say. If this is so, the history of language is a history not of animal communication, but of animal systems of representation. More broadly, the history of language is the history of consciousness.

By representation Bickerton means simply "...responding or having a permanent propensity to respond to *x*, an entity or event in the external world, in terms of *y*, a particular pattern of neural activity" (p. 76). On this definition, every creature with a nervous system, from earthworm to human, has a representation of the world. Representations are normally veridical (if they were not, how would the creature survive?), but incomplete, adequate to their function, but not optimal. Each species has evolved sensory systems to pick up information about the world necessary for survival, motor systems to respond to that information, and representational systems to link information with behavior. Thus, each species carves reality into a set of functional categories that constitutes its characteristic "primary representational system (PRS)" (p. 82).

PRSs can be simple or complex. Natural selection has favored the emergence of increasingly complex systems, not because prior systems were inadequate (cockroaches and frogs are still with us, and thriving), but because sensitivity to more and subtler aspects of the environment permits an animal to gain an advantage over its fellows, to leave more offspring, and ultimately perhaps to found a new species. Thus, there is a true "...*scala naturae*...based on the degree to which creatures are capable of representing the world around them" (p. 102).

By the time the more advanced primates arrived on the scene, much of the infrastructure for language, a secondary representational system (SRS) grafted onto the first, had been laid down. All that was needed was the appropriate selection pressures to "...make the development of secondary representations advantageous, and some means through which such representations could be made both concrete and communally available" (p. 101). Notice that, on this view, evolution has a direction (not a popular idea among evolutionists) toward increasingly comprehensive systems of representation. Language then lies "...in the mainstream of evolution...merely a single (if admittedly rather a large) step in an orderly process, the development of more and more sophisticated ways of representing (that is, knowing) a world external to the creature" (p. 102). This is how Bickerton resolves the Continuity Paradox.

In what sense, then, is language a system of representation? Bickerton adopts the metaphors of a map, and of a handbook of itineraries: the map is the lexicon, a chart of semantic space, the itineraries are the sentences, an infinity of paths through that space. At first, we must presume, symbols on the lexical map (that is, words) referred only to categories, or concepts, in the human PRS, corresponding to objects and events in the external world—as they still do for the child first learning to talk. But because words refer to concepts of the world, not to the world itself, language was able to evolve into an autonomous system of representation, deploying words that refer to abstract concepts for which there are no perceptual correlates, and even to imaginary objects that do not exist. Thus, words mirror not reality, but a speaker's, a culture's, a species' view of reality.

If we want to clear a path through semantic space, that is, if we want to represent unambiguously 'who did what with which to whom,' we cannot simply string words together in any sequence we like. Not all paths across a terrain are passable, and not all strings of words form sentences. Permissible strings are shaped by structural principles (predicability, grammaticization, syntax) of which the most important, and the one that "...most decisively separates us from other species" (p. 57), is syntax. In a scant, but lucid, fifteen pages Bickerton sketches the descriptive principles of the X-bar theory of phrase structure, "...the very core of universal syntax" (p. 61), the "universal" (p. 67) argument structure of verbs, and the principles by which the latter is mapped onto the former according to a hierarchy of thematic roles. He emphasizes that talkers and listeners put their syntactic machines to work automatically and unconsciously, and he concludes that "...we are in the presence, not merely of a system that determines what we can and cannot say, but of a system wholly below the level of consciousness that cannot have been acquired through instruction, induction, or any of the other recognized processes by which learning takes place" (p. 73).

With the division of language into two autonomous components, lexicon and syntax, Bickerton effectively also divides the question of language origins into two distinct questions: (1) How did arbitrary elements acquire reference? (2) How did syntax develop? For the first step "all that was necessary was for some kind of label to be attached to a small number of preexisting concepts" (p. 128). Evidence for this step comes from "the fossils of language" (Chapter 5). Here Bickerton turns on its head the argument that, since the utterances of a signing chimpanzee and a talking two year old child are formally identical, the chimpanzee is at least learning the rudiments of language. On the contrary, Bickerton responds, neither chimpanzee nor child is learning language. Instead, both are using "protolanguage," a phylogenetic precursor of true language that is recapitulated in the child and can be elicited by training from the chimpanzee, because an SRS is already latent in its well developed conceptual representation of the world.

Speakers (or signers) of a protolanguage have a referential lexicon but, with the possible exception of some negators, question words and quantifiers, no grammatical items and no syntax. Specifically, their utterances display no regularities of word order, no systematic expansion of structure in phrases and clauses, no obligatory expression of subcategorized arguments, and no automatic identification of null elements. Nonetheless, we are justified in regarding protolanguage as a unitary mode of representation, peculiar to our species, because it emerges naturally by mere exposure to words not only in children under two, but in older children deprived of language during the "critical period" (e.g., Genie, the well-known "wild child" of California), and even in adults obliged to communicate in a second language of which they know only a few words. This last was the fate of entire communities of slaves, speakers of diverse native languages, who were thrown together in European colonies of the Caribbean and Pacific. Forced to communicate in the language of their masters, but denied (due to the adverse ratio of learners to speakers) the opportunity to learn that language fully, they developed "pidgins," formally identical, Bickerton tells us, with the protolanguage of under-twos.

Now if chimpanzees and an indefinite number of other species do indeed possess a latent SRS so close to the surface that it can be elicited by laboratory conditioning, why did an SRS evolve only in the hominid line? The answer, Bickerton speculates, lies in the conditions of life on the savanna. There, lacking the steady food supply of the forests, hominids were obliged to hunt, gather and scavenge. Lacking too the speed, strength, teeth and claws of rival predators (to which they themselves might fall prey), hominids had to live by their wits. These conditions set a premium on continued evolutionary sharpening of such primate skills as stereoscopic vision, bipedal locomotion (affording a wide field of view while tracking game) and manual skills, including tool use. More broadly, the conditions set a premium on the evolution of curiosity, close observation, long-term memory and a generally enriched PRS. Eventually, a point was reached at which any further gain in knowledge of the world could only come from the emergence of an SRS.

We shall probably never know when hominids first stumbled on the discovery that arbitrary grunts and mouthings could be associated with an absent referent, and the exact date is not in any case critical to the argument. But from a careful consideration of the palaeoanthropological evidence Bickerton concludes that protolanguage probably emerged in *Homo erectus* (c. 1.5 - 0.5 million years ago (mya)). This would be consistent with the substantially higher brain/body ratio of *Homo erectus* than of *Homo habilis* (c. 2.5 - 1.5 mya), and with the substantially lower brain/body ratio of *Homo erectus* than of *Homo sapiens*, the initiator of true language.

The advantages of being able to communicate in a protolanguage, even if its lexicon is modest, are not difficult to imagine. However, consistent with his view of language as primarily a system of representation, Bickerton emphasizes not communication, but "the way in which protolanguage would have remodeled...[the hominid's]...internal world" (p. 156). The greatest conceptual gain would perhaps be in learning. Here Bickerton distinguishes among three modes of learning: experiential, observational, and what he terms "constructional." A creature capable of constructional learning can increase its knowledge of the world by forming propositions about objects, events, behaviors that are not immediately present to its senses. For example, such a creature might have observed that a certain kind of predator tends to avoid water. The creature could then "...construct a kind of propositional machine:

'Predator avoids water. Suppose predator chases me. I cross water. Predator does not cross water. I escape.' Thus despite the absence of any direct example, the first time the creature is pursued by that predator, it escapes by swimming" (p. 159).

Virtually all animals can learn by experience (for which operant conditioning is the paradigm), and many species can learn by observation. But few (perhaps only one) can learn by constructing propositions. Certainly, chimpanzees can solve novel problems by stacking boxes or using a stick to reach otherwise inaccessible food, but seemingly they can do so only if all the elements needed for the solution are present in the immediate environment. Arguably, then, only a creature with an SRS can construct an abstract model of possible behaviors with respect to absent objects, and try them out in its mind to see if they work. The advantages that will accrue to such a creature are obvious. Bickerton indeed proposes that the large increase in brain/body ratio during the million or so years from *erectus* to *sapiens* was mainly due to a continuous feedback process between constructional learning, made possible by the evolving protolanguage, and the need for neural space to store an increasingly subtle conceptual representation of the world.

Nonetheless, the protolanguage of *erectus* had its limits. There is no evidence from the fossil record that *erectus* was the cognitive equal of *sapiens*. One limit was perhaps on the size of the lexicon due to the relatively small number of discriminably different words that *erectus* could articulate with its still not fully developed vocal tract. A second limit was set by the lack of grammatical items (verbal inflections, auxiliaries, prepositions) for representing the time, locus and direction of objects, actions and events. The third and most serious limit was on thinking itself, due to the lack of syntax. For human thinking, Bickerton argues, is intrinsically syntactic, "...since it depends crucially on the existence of structures like: *x* happened because *y* happened; whenever *x* happens, *y* happens; unless *x* happens, *y* will not happen; if *x* happens, *y* or *z* will happen; although *x* happened, *y* did not happen; *x* thinks, believes, says, hopes, fears, knows that *y* will or will not happen; *x* does *y* in order to *z*; *x* wants the *y* that *z*-ed, not the *a* that *b*-ed; and so on" (p. 162).

To *Homo sapiens* then was left the great leap into syntax. That it was indeed a leap, a saltation not a gradual development, is the final step in Bickerton's argument. He adduces three lines of evidence against gradualism. First is ontogenetic evidence, from two areas, that protolanguage can change into true language without an intervening stage. In a typical normal child, Bickerton argues from transcripts of a certain child's utterances at 21 months and 27 months, at least four (perhaps all) of the five properties noted above as absent from protolanguage come in as a cluster within a 6-month period (grammatical items, systematic expansion of phrase structure, obligatory expression of subcategorized arguments, automatic identification of null elements and, perhaps, systematic variation of word order). All five properties also appear in creole languages, the languages spoken by first generations of children raised among pidgin-speaking adults. If these children were indeed exposed only to protolanguage, as seems to have been the case, their language is a "...direct expression of a species-specific biological characteristic, a capacity to recreate language in the absence of any specific model from which the properties of language could be 'learned' in the ways we normally learn things" (p. 171). The ontogenetic evidence then demonstrates that a leap from protolanguage to language is possible.

A second line of evidence (or at least argument) against gradualism demonstrates, on linguistic grounds, that a leap is not only possible, but necessary. Any proposal for an "interlanguage" is likely to entail either a partial application of full language rules or a contravention of them. As far as partial applications go, there is no reason to suppose that it is easier to apply a rule in some contexts, but not in others, than to apply it across the board. And as for contraventions, such as mapping thematic roles directly onto surface ordering (e.g., all grammatical subjects are Agents, all objects are Patients), this would require 'unlearning' the interlanguage rules to make the step into language—a biologically implausible detour on the road from protolanguage to language.

A final line of evidence against gradualism comes from recent studies of mitochondrial DNA (mtDNA) in modern populations. MtDNA molecules are inherited maternally through the oocyte and, unlike nuclear genes, cannot be lost through recombination. Studies tracing the history of mtDNA molecules have given rise to the so-called African Eve hypothesis that *Homo sapiens* originated as a distinct species due to (in Bickerton's words) "...an event, presumably a mutation of some kind, that affected a single female living in Africa... between 140 and 290 Kya (thousand years ago)" (p. 165). The hypothesis of a single, common ancestor is widely, though not universally, accepted by both geneticists and palaeoanthropologists. The issue is a technical one that, as Bickerton remarks, must be decided by professionals in the field. Nonetheless, he himself adopts the common ancestor hypothesis, terming it the "catastrophic" view, largely on the basis of the fossil record. This record shows remarkably little change in the hominid "tool kit" over the million or so years of the *erectus* period. But around 70 Kya, soon after the appearance of *sapiens*, the record is suddenly enriched by bladed tools, cave paintings, stone figurines, moon calendars—a panoply of artifacts that, Bickerton argues, could only have been produced by a creature with the powers of thought and constructional learning that language affords.

Bickerton is aware of the difficulties that a catastrophic account brings in its train, and he devotes many pages to puzzling them through. For example, he lists a minimal set of eight, "quasi-independent entities" (p. 189) or components in the current generative model from which all the subtle phenomena of descriptive syntax (including the five criteria of true language, noted above, and many more) might be deduced: hierarchical structure (X-bar theory), binding, control, bounding, theta theory, conditions for government, conditions for case-assignment, and movement processes. He considers how these components might be collapsed into a smaller number, thereby reducing the number of required mutations. But even if we managed to reduce the eight "entities" to a single principle, he believes that we would still not be out of the woods, because "...our species is distinguished from all others not merely by syntacticized language, but also by changes in the features and dimensions of the skull and by our typical supralaryngeal tract. How could any single genetic event...occasion so many and such diverse changes?" (p. 190). In the end, after much intellectual writhing, Bickerton pins his hopes for a solution on advances in our understanding of how brains are organized and how they develop.

With language, then, came the only forms of thought, mind and consciousness that we as a species can know. In the last two chapters of his long argument, Bickerton explores some philosophical implications and presumed historical consequences of the assumptions that language is the essential mode of human thought and the source of our self-reflexive consciousness. Only by understanding language and its evolution, Bickerton concludes, will we understand ourselves and our place in nature.

The most interesting and original aspect of this book is also the most speculative: the formulation and resolution of the Continuity Paradox. Discussions of language origins seldom acknowledge the vast infrastructure of knowledge and beliefs about the world that underlies even a simple spoken utterance. No other writer that I have read places the issue, where it belongs, at the center of an evolutionary account of language. Of course, any story about the origins of "mind" will be speculative, but I found Bickerton's notion of increasingly complex systems of primary representation, from which language emerges as a secondary system, persuasive and intuitively appealing. He argues his case for the continuity of animal mind and human mind clearly, subtly and with great erudition.

However, if we ask how successful Bickerton is in reconciling linguistic theory with neo-Darwinian theory, the answer is, Not very. The broad reason for this is that he forces all the concessions out of biology, none out of linguistics. Given the central, unifying role of evolutionary theory in modern biology, and the uncertain, still developing role of generative grammar in linguistics, this is the wrong way round. More narrowly, there are, in my view, two main flaws in Bickerton's argument. First is his rejection of gradual, step-by-step evolution in favor of "catastrophic" saltation. Second, is his disregard of the communicative aspects of language function by which he shuts himself off from explaining many aspects of linguistic form.

In Darwin's (1859/1964) often quoted discussion of "organs of extreme perfection and complication" (of which we can take the neural underpinnings of language to be a supreme example), he remarks: "If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down" (p. 189). For many years Darwin's claim was the target of attack from proponents of alternative "saltationist" accounts, but the development of Mendelian genetics in the first half of the 20th century made clear that the only way complex structures finely adapted to their function can evolve is through a long sequence of mutations with small effects. By the centenary of the publication of the *Origin* it was possible for Mayr (1960), the greatest living authority on evolutionary theory, to write:

> The genetic work of the last four decades has refuted mutationism (saltationism) so thoroughly that it is not necessary to repeat once more all the genetic evidence against it. Most important is that the phenotype (in higher organisms) is the product of a long developmental pathway and ...any part of it, any 'character,' depends on the harmonious interaction of many, if not all, the genes of the organism (p. 356).

Bickerton is certainly aware of the established dogma. Indeed, in his earlier book, *Roots of Language* (1981), he accepted it as part of the "biological reality" (1981, p. 294) to which any serious theory of language and language origins has to conform. He was at pains to chart a plausible step-by-step course for the possible evolution of a stripped-down innate core of language structure, the "bioprogram," through lexicalization of the presumed primate semantic space, the emergence of thematic roles implicit in verbs, grammaticization of lexical items, the development of recursion from the serial stringing of verbs, and so on (1981, Chapter 4). In that book too he poured scorn on those who believe that language sprang "...in its entirety from Jove's brow by some beneficent and unprecedented mutation" (1981, p. 215), and he stressed that "...evolution has advanced not by leaps and bounds but by infinitesimal gradations" (1981, p. 221).

Why then does he now baldly, without acknowledging the eccentricity of his step, adopt a saltationist account of language origins and allow himself to become mired beyond rescue in the attempt to extricate himself from the consequences? From internal evidence, as scholiasts say, the answer seems to be that in the interval between the books two things happened to change Bickerton's mind. First, the influence of Chomsky, acknowledged in the preface as "long resisted" (p. x), came to bear; the change is signaled by a switch from the vigorous, even truculent, polemic with which he attacked Chomsky's work in the first book to the tone of chastened respect adopted in the second. For no doubt sound reasons that he unfortunately doesn't tell us, Bickerton abandoned his language bioprogram hypothesis (Bickerton, 1984) and adopted in its place the "universal grammar" of current generative theory.

Now, the formal apparatus of generative grammar is purely descriptive, a matter to be explained rather than itself explanatory. Nothing intrinsic to its principles demands a home-made biology to account for them. Yet, for reasons perhaps having more to do with protecting the autonomy of their field than with science, generative theorists find the only acceptable "explanation" to be that the principles are "innate." Attempts to understand how they might have become innate in the first place, that is, to trace their functional origins in evolution and development, are dismissed out of hand (see, for example, many of the commentaries on Pinker and Bloom (1990)). But why did Bickerton, hitherto a sturdy functionalist, fall into line? My guess is that he finally recognized (what was obvious to a reader of the first volume) that the pin of his entire argument, the claim that children raised among pidgin speakers invented creole languages in a single generation, was no more than an extreme statement of the familiar argument from "the poverty of the stimulus" combined with "...the Chomskyan idealization of 'instantaneous acquisition' " that he had once ridiculed (Bickerton, 1981, p. 142)

In any event, having bartered functionalism for descriptive power, Bickerton was momentarily beached, caught between the entirely orthodox evolutionary functionalism he had adopted to resolve the paradox of the continuity of human and animal mind, and the flagrant anti-functionalism of generative grammar. Then the second thing happened. In 1987 Chance lowered onto the stage the *deus ex machina* of the African Eve hypothesis, granting (Bickerton evidently believes) a new license to saltationism. He interprets the discovery of a possible common ancestor for all living humans as evidence that *Homo sapiens* emerged full-blown, with all its diverse species-specific characters, from a single "crucial mutation" (p. 190). Yet the researchers who first proposed the hypothesis of a common ancestor, in a paper that Bickerton cites, explicitly deny the necessity of this interpretation:

> ...our placement of the common ancestor of all human mtDNA diversity in Africa 140,000 - 280,000 years ago need not imply that the transformation to anatomically modern *Homo sapiens* occurred in Africa at this time. The mtDNA data tell us nothing of the contributions to this transformation by the genetic and cultural traits of males and females whose mtDNA became extinct (Cann, Stoneking, & Wilson, 1987, p. 35).

Accordingly, even if we accept the Eve hypothesis, we are not obliged to believe that evolution stopped with her birth. Much can have happened during the 210,000 ± 70,000 years between the hypothesized point of the species' origin and its radiation out of Africa, marked by the relatively sudden appearance of novel artifacts (roughly 70 Kya). If we accept, further, Bickerton's argument that syntax began with *Homo sapiens*, we would then have, at four generations to a century, between roughly 3,000 and 8,000 generations for its basic principles to evolve. Whether this span would have been enough to do the job, I don't know. Certainly, it does not seem so short that we should be pressed into abandoning a basic tenet of evolutionary theory.

However, Bickerton's flight from gradualism was abetted by a second closely related flaw in his argument that springs from a misconception of how genes relate to the structures they help to form. The misconception is evident, for example, in his notion that, if we could collapse the eight quasi-independent components of the current generative model to, say, three or four, we would similarly reduce the number of mutations required to account for them. The objection to this notion, implicit in the quotation from Mayr (1960) above, is that there is no one-to-one correspondence between genes and phenotypic characters: "Mendelism is a theory of particulate inheritance, not of particulate embryology" (Dawkins, 1982, p. 116)—and still less, we may add, of particulate behavior, or particulate structures underlying behavior. A single phenotypic character may depend on the action of many genes, and a single gene may contribute to the formation of many characters. If we recall further that some forty genes work together to form the wing of a fruit fly (Waddington, 1975, p. 72), it seems wildly implausible that one or even one hundred genes suffice to encode the (perhaps) specialized neural structures underlying human language capacity.

What all this means is that the goal of mapping language onto genes is, at present and perhaps indefinitely, scarcely intelligible. For example, Gopnik (1990) recently reported the results of a pedigree analysis of an extended family consistent with the hypothesis that a certain form of dysphasia, which she termed "feature blindness," arose from a defect in a single dominant gene. Let us suppose that further analysis of this family were to isolate the chromosome, and even the precise stretch of DNA on the chromosome, where the defect lies. We would then have our hands on *a* gene, but not on *the* gene, for feature sensitivity. The reason for this is simply that many other genes are likely to be no less crucial for the development of normal sensitivity to features than the one we have isolated. In fact, if we were to conduct a similar study of another family, in which there ran the same form of dysphasia, we would be quite likely to end up isolating a different gene, perhaps even on a different chromosome. In short, neither genetic fantasy nor genetic fact can restore Eve to the saltationist Eden from which genetic knowledge drove her out.

The reader who has happily followed Bickerton in his resolution of the first Continuity Paradox may be taken aback by his blithe introduction of a second Continuity Paradox to "solve" the puzzle of syntax. Nonetheless, the story hangs together. If we add to the new Discontinuity a defensible (though, in this book, weakly defended (p. 115)) assumption that language development recapitulates, within limits, language evolution, many otherwise puzzling aspects of Bickerton's argument fall into place: the sharp division of lexicon from syntax, the notion of an encapsulated protolanguage distinct from true language, the denial of a language mode intermediate between the two, the supposed ontogenetic "leap" (albeit of 6 months duration) by a normal child into syntax, the uncritical acceptance of a "critical period" for language acquisition, and not least the assumption that language is primarily a system for representing, rather than communicating, propositions.

This last assumption, a crucial first line of defense against functionalism, oversimplifies the relation between language and thought. We can certainly think without speaking, but we also often speak to find out what we think. In other words, communication facilitates thought no less than the reverse—as indeed Bickerton acknowledges in the preface: "How can you know what you think until you've disagreed with someone?" (p. x). In fact, Bickerton concedes that: "Communication...was perhaps the only means by which either [language or protolanguage] could have bootstrapped its way up from mere latency to the status of a useable representational system" (p. 146). But instead of treating the communicative and representational functions as mutually reinforcing components of a feedback system—the more you say, the more you have to say, and vice versa—he disregards the communicative function almost entirely. The result is that he seriously weakens his account of the origins of language form because he shuts off explanations of properties that presuppose a listener—most notably, the rules of phonetics and phonology instantiated in "articulate speech."

Curiously, although phonology is touched on only three times in the book (pp. 14, 144-5, 162) and earns no entry in the index, Bickerton does recognize "...commonalties between phonology and syntax which suggest that the evolution of the two may have been linked" (p. 144). He doesn't say what these commonalties are. But if they are evolutionarily based, they must extend beyond mere formal analogy at least to the types of selection pressures that shaped them, and perhaps even to homologies in their neural substrates. With regard to the pressures, Bickerton goes so far as to suggest that the words of the first hominid talkers may have been "unanalyzable wholes" and that as vocabulary grew, it would have exerted a selective pressure toward reliable articulation and comprehension (p. 145). Here he is on the edge of sketching an account of the origin of sound systems along the lines of the only quantitative model so far proposed (Lindblom, 1988, and references therein). This model predicts the phonemic elements in a system of sound contrasts from competing selection pressures toward ease of comprehension and ease of articulation, within the presumed constraints of auditory and articulatory capacity. While we still have everything to learn about possible non-linguistic neurobehavioral constraints on syntactic form, we have no reason to suppose that pressures toward ease and clarity of expression ceased to operate just for the evolution of syntax.

Had Bickerton pursued such a functionalist line of thought he might have been led toward more serious consideration of the social uses of language, and even toward reflections on the possible social, rather than purely linguistic, origins of a self-reflexive consciousness (De Waal, 1984; Humphrey, 1983). On such a view, as consciousness of self, and so of other selves, developed, new words with their implicit syntactic requirements, would have sprung up to capture the increasingly subtle concepts and feelings of individual and social life (cf. p. 187). From there the path into a grammar projected by the lexicon, —explored in Bickerton's first book and briefly scouted again in this one (pp. 181-188)—would have been straight. No doubt, such a lexicalist account of syntax would have left much to be explained, but surely a good deal less than we are left with here. Of course, that would have been another book by another writer—perhaps, indeed, by Bickerton's former self!

I have said enough about the book's weaknesses. Let me end with its strengths. Perhaps there are those, like Chomsky himself, who view questions about the origins of language as unanswerable and therefore of no more interest than questions about the origin of the heart or the arm (see Chomsky's remarks in Harnad, Steklis and Lancaster, 1976, p. 57). But more than a century of evolutionary studies have demonstrated that such questions are not necessarily unanswerable. Moreover, there are many who consider them to be of quite extraordinary interest because in trying to answer them we throw light on the structure and function of the organs examined. This is a book for such people. Bickerton calls it "...a machine for thinking about language and what language has done for our species and how it has made us different from other species" (p. 6). The book commands a large and diverse range of knowledge, constantly raises unexpected questions leading into new lines of thought, and adjudicates every issue with an admirable, even-handed rationality. Few readers will come away without a deepened understanding of the scope and limits of human language. I believe that Thomas Huxley, concerned as he was with where humans fit into the scheme of things, would have been pleased to read it, and even more pleased to have written it.

REFERENCES

Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma Publishers, Inc.

Bickerton, D. (1984). The language bioprogram hypothesis. *The Behavioral and Brain Sciences, 7,* 173-221.

Bickerton, D. (1990). *Language and species*. Chicago: University of Chicago Press.

Cann, R. L., Stoneking, M., & Wilson, A. C . (1987). Mitochondrial DNA and human evolution. *Nature, 325,* 31-36.

Darwin, C. (1859/1964). *On the origin of species*. Cambridge, MA: Harvard University Press. (Facsimile of the first edition).

Dawkins, R. (1982). *The extended phenotype*. New York: Oxford University Press.

De Waal, F. (1984). *Chimpanzee politics*. New York: Harper & Row.

Gopnik, M. (1990). Feature-blind grammar and dysphasia. *Nature, 344,* 715.

Harnad, S. R., Steklis, H. D., & Lancaster, J. (1976). *Origins and evolution of language and speech*. New York: Annals of the New York Academy of Sciences, Vol. 280.

Humphrey, N. (1983). *Consciousness regained*. New York: Oxford University Press.

Huxley, T. H. (1863/1915). *Man's place in nature*. New York: Appleton.

Lindblom, B. (1988). Models of phonetic variation and selection. In L. Cavalli-Sforza & A. Piazza (Ed.), *Language change and biological evolution*. Stanford, CA: Stanford University Press.

Mayr, E. (1960). The emergence of evolutionary novelties. In S. Tax (Ed.), *Evolution after Darwin. I. The evolution of life: Its origin, history, and future* (pp. 349-380). Chicago: University of Chicago Press.

Pinker, S., & Bloom, P. (1990) Natural language and natural selection. *Behavioral & Brain Sciences, 13,* 707-784.

Waddington, C. H. (1975). *The evolution of an evolutionist*. Chicago: University of Chicago Press.

FOOTNOTES

[*] *Applied Psycholinguistics*, in press.

[†] Also University of Connecticut and Yale University.

Appendix

SR #	Report Date	DTIC #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 071-484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A055853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 196-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-90
SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	PB 88-244256	ED 282-278

SR-89/90	January-June 1987	PB 88-244314	ED 285-228
SR-91	July-September 1987	AD A192081	**
SR-92	October-December 1987	PB 88-246798	**
SR-93/94	January-June 1988	PB 89-108765	**
SR-95/96	July-December 1988	PB 89-155329/AS	
SR-97/98	January-July 1989	PB 90-121161/AS	ED321317
SR-99/100	July-December 1989	PB 90-226143/AS	ED321318
SR-101/102	January-June 1990	PB 91-138479	
SR-103/104	July-December 1990	PB 91-172924	ED331100
SR-105/106	January-June 1991	PB92-105204	
SR-107/108	July-December 1991		

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned